

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2002, 01:04:31 : Search time 62.27 Seconds

(without alignments)  
242.589 Million cell updates/sec

Title: US-09-728-670-10

Perfect score: 715

Sequence: 1 SSSFDKGRKKKGDASYFEP.....SEHIKNPGFNLITKVIEKK 136

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	715	100.0	136	17	AAW03071	Wild-type staphylo
2	715	100.0	136	17	AAW03086	Wild-type staphylo
3	715	100.0	136	22	AAB99421	Staphylokinase (Sa
4	715	100.0	136	22	AAB61909	S. aureus staphylo
5	715	100.0	137	14	AAR39150	Staphylokinase SAK
6	712	99.6	136	20	AAV15024	Staphylokinase var
7	711	99.4	136	19	AAW44696	Staphylokinase fro
8	711	99.4	136	20	AAV15022	Staphylokinase var
9	711	99.4	136	20	AAV15025	Staphylokinase var
10	711	99.4	136	20	AAV15016	Staphylokinase var
11	711	99.4	136	20	AAV15021	Staphylokinase var

12	711	99.4	136	22	AAB61908	S. aureus staphylo
13	711	99.4	137	14	AAR39149	Staphylokinase SAK
14	711	99.4	163	12	AAR12137	S. aureus Staphylo
15	711	99.4	163	13	AAR28844	Staphylokinas (SAK
16	710	99.3	136	20	AAV15013	Staphylokinase var
17	709	99.2	136	17	AAW03101	Staphylokinase der
18	709	99.2	136	20	AAV15023	Staphylokinase var
19	709	99.2	136	20	AAV15015	Staphylokinase var
20	709	99.2	136	20	AAV15010	Staphylokinase var
21	709	99.2	136	20	AAV15004	Staphylokinase var
22	709	99.2	136	20	AAV15006	Staphylokinase var
23	708	99.0	136	19	AAW44690	Staphylokinase mut
24	708	99.0	136	19	AAW44693	Staphylokinase mut
25	708	99.0	136	20	AAV15026	Staphylokinase var
26	708	99.0	136	20	AAV15027	Staphylokinase var
27	707	98.9	136	19	AAW44694	Staphylokinase var
28	707	98.9	136	20	AAV15007	Staphylokinase mut
29	707	98.9	136	20	AAV15018	Staphylokinase var
30	707	98.9	136	20	AAV15019	Staphylokinase var
31	707	98.9	136	20	AAV15011	Staphylokinase var
32	706	98.7	136	19	AAW44689	Staphylokinase mut
33	706	98.7	136	19	AAW44692	Staphylokinase mut
34	706	98.7	136	20	AAV15028	Staphylokinase var
35	706	98.7	136	20	AAV15014	Staphylokinase var
36	706	98.7	136	21	AAB01297	Wild type staphylo
37	705	98.6	136	19	AAW44691	Staphylokinase mut
38	705	98.6	136	19	AAW44695	Staphylokinase mut
39	704	98.5	136	20	AAV15041	Staphylokinase var
40	704	98.5	136	20	AAV15012	Staphylokinase var
41	704	98.5	136	20	AAV15012	Staphylokinase var
42	703	98.3	136	17	AAW03079	Staphylokinase der
43	703	98.3	136	17	AAW03078	Staphylokinase der
44	703	98.3	136	17	AAW03082	Staphylokinase der
45	703	98.3	136	17	AAW03076	Staphylokinase der

#### ALIGNMENTS

RESULT	ID	AAW03071	standard; protein; 136 AA.
XX	AC	AAW03071;	
XX	DT	19-FEB-1997 (first entry)	
XX	DE	Wild-type staphylokinase.	
XX	KW	Staphylokinase; mutant; mutein; variant; immunogenicity; decrease; derivative; SakSTAR; arterial thrombosis; thrombolytic agent.	
OS	XX	Staphylococcus aureus.	
FH	Key	Location/Qualifiers	
FT	Region	5..6	
FT	Region	/label= M20_epitope	
FT	Region	8..10	
FT	Region	/label= M21_epitope	
FT	Region	11..14	
FT	Region	/label= M1_epitope	
FT	Region	19	
FT	Region	/label= M22_epitope	
FT	Region	33..35	
FT	Region	/label= M2_epitope	
FT	Region	35..38	
FT	Region	/label= M3_epitope	
FT	Region	46..50	
FT	Region	/label= M4_epitope	
FT	Region	57..59	
FT	Region	/label= M5_epitope	
FT	Region	61..65	
FT	Region	/label= M6_epitope	

FT Region 65..69  
FT /label= M7\_epitope  
FT Region 74..77  
FT /label= M8\_epitope  
FT Region 80..82  
FT /label= M9\_epitope  
FT Region 86..88  
FT /label= M10\_epitope  
FT Region 93..94  
FT /label= M11\_epitope  
FT Region 96..98  
FT /label= M12\_epitope  
FT Region 99..100  
FT /label= M13\_epitope  
FT Region 99..102  
FT /label= M14\_epitope  
FT Region 108..109  
FT /label= M15\_epitope  
FT Region 115..119  
FT /label= M16\_epitope  
FT Region 119..121  
FT /label= M17\_epitope  
FT Region 130  
FT /label= M18\_epitope  
FT Region 134..136  
FT /label= M19\_epitope

EP721982-A1.

17-JUL-1996.

06-JAN-1995; 95EP-0200023.

06-JAN-1995; 95EP-0200023.

(COLL/) COLLEN D J.

(LEUV-) LEUVEN RES &amp; DEV VZW.

Collen D;

WPI; 1996-322832/33.

New staphylokinase mutants with reduced immunogenicity - useful for treating arterial thrombosis

Disclosure; Page 16; 21pp; English.

CC staphylokinase derivatives showing a reduced immunogenicity as compared to wild-type staphylokinase are claimed. The derivatives are useful as thrombolytic agents to treat arterial thrombosis and are pref. produced by eliminating at least one of the epitopes indicated in the features table. The epitopes are destroyed by replacing one or more amino acid residues in a charge cluster by an Ala residue. Mutations are introduced using site-directed mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus lysogenic strain SakSTAR. The present sequence is that of wild-type SakSTAR staphylokinase.

SQ Sequence 136 AA;

Query Match 100.0%; Score 715; DB 17; Length 136;  
Best Local Similarity 100.0%; Pred. No. 1.9e-70;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSSFDKGYKKGDDASYFEPTGPLYMVNTGVDSKGNELISPHYVEFPKPGTTLTKRKI 60  
Db 1 sssfdkgykkgddasyfepgplymvntgvdsksgnelispkyvefpikpgtltltekki 60  
OY 61 EYVYEWALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSEPTTEKGFVVPDSEHI 120  
Db 61 eyvewaldataykefrvvelpsakievtydyknkkkeetksfptekgfvpdlsehl 120

OY 121 KNPGFNLTFRVIEKK 136  
Db 121 knpgfnlfrkviekk 136

RESULT 2

AAW03086 standard; protein; 136 AA.

AC AAW03086;

DT 19-FEB-1997 (first entry)

DE Wild-type staphylokinase.

KW Staphylokinase; mutant; variant; immunogenicity; decrease;  
KW derivative; SakSTAR; arterial thrombosis; thrombolytic agent.

Staphylococcus aureus.

Location/Qualifiers

FT Region 5..6  
FT /label= M20\_epitopeFT Region 8..10  
FT /label= M21\_epitopeFT Region 11..14  
FT /label= M1\_epitopeFT Region 19  
FT /label= M22\_epitopeFT Region 33..35  
FT /label= M2\_epitopeFT Region 35..38  
FT /label= M3\_epitopeFT Region 46..50  
FT /label= M4\_epitopeFT Region 57..59  
FT /label= M5\_epitopeFT Region 61..65  
FT /label= M6\_epitopeFT Region 65..69  
FT /label= M7\_epitopeFT Region 74..77  
FT /label= M8\_epitopeFT Region 80..82  
FT /label= M9\_epitopeFT Region 86..88  
FT /label= M10\_epitopeFT Region 93..94  
FT /label= M11\_epitopeFT Region 96..98  
FT /label= M12\_epitopeFT Region 99..100  
FT /label= M13\_epitopeFT Region 99..102  
FT /label= M14\_epitopeFT Region 108..109  
FT /label= M15\_epitopeFT Region 115..119  
FT /label= M16\_epitopeFT Region 119..121  
FT /label= M17\_epitopeFT Region 130  
FT /label= M18\_epitopeFT Region 134..136  
FT /label= M19\_epitope

PN WO9621016-A2.

PD 11-JUL-1996.

PF 03-JAN-1996; 96WO-EP00081.

PR 17-NOV-1995; 95JP-0299781.

PR 06-JAN-1995; 95EP-0200023;  
PR 11-JAN-1995; 95US-0371505.  
PR 09-JUN-1995; 95EP-0201531.  
PR 06-JUL-1995; 95US-0499092.  
XX  
PA (COLL-) COLLEN D.  
PA (LEUV-) LEUVEN RES & DEV VZW.  
XX  
PI Collen D;  
XX  
DR WPI; 1996-333991/33.  
XX  
PT New staphylokinase derivs. having reduced immunogenicity - useful  
PT for treating arterial thrombosis  
XX  
PS Disclosure; Fig 1; 58pp; English.  
XX  
CC Staphylokinase derivatives showing a reduced immunogenicity as  
CC compared to wild-type staphylokinase are claimed. The derivatives  
CC are useful as thrombolytic agents to treat arterial thrombosis and  
CC are pref. produced by eliminating at least one of the epitopes  
CC indicated in the features table. The epitopes are destroyed by  
CC replacing one or more amino acid residues in a charge cluster by an  
CC Ala residue. Mutations are introduced using site-directed  
CC mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus  
CC lysogenic strain SakSTAR. The present sequence is that of wild-type  
CC SakSTAR staphylokinase.  
XX  
SQ Sequence 136 AA;  
XX

Query Match 100.0%; Score 715; DB 17; Length 136;  
Best Local Similarity 100.0%; Pred. No. 1.9e-70;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 SSSFDDGKYYKKGDASFEPTGPYLMVNTGVDSKGNELSPHYVEFPPIKPGTTLTKEKI 60  
Db 1 sssfdkgkykkgdasyfeptgpylmvntgvdsksgnellsphylvefpikpgtltlkeki 60  
XX  
QY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
Db 61 eyvewaldataykefrvvelpsakievtyydknkkkeetskspitekgfvvpdlsehi 120  
XX  
QY 121 KNPGFNLTIKVIEKK 136  
Db 121 knpgfnltikvlekk 136  
XX

RESULT 3  
AAB99421  
ID AAB99421 standard; Protein; 136 AA.  
XX  
AC AAB99421;  
XX  
DT 28-AUG-2001 (first entry)  
XX  
DE Staphylokinase (SakSTAR) primary protein sequence.  
XX  
KW Staphylokinase; Staphylococcus aureus; SakSTAR; T-cell epitope;  
KW immunogenic; thrombolytic; acute myocardial infarction; immunogenicity.  
XX  
OS Staphylococcus aureus.  
XX  
PN WO200140281-A2.  
XX  
PD 07-JUN-2001.  
XX  
PF 01-DEC-2000; 2000WO-EP12299.  
XX  
PR 02-DEC-1999; 99EP-0204093.  
XX  
PA (THRO-) THROMB-X NV.  
XX

PI Warmerdam PAM, Plaisance SONGH, Collen DJ, De Maeyer MCH;  
XX  
DR WPI; 2001-374786/39.  
XX  
PT Reducing immunogenicity of protein (P) by eliminating T cell epitopes  
PT in test peptides having amino acid sequence corresponding to (P) and  
PT modifying amino acid sequence of (P) according to test peptide  
PT modifications -  
XX  
PS Disclosure; Fig 1; 50pp; English.  
XX  
CC The present invention describes a method for reducing the immunogenicity  
CC of a peptide or protein (I), where (I) can be staphylokinase (SakSTAR)  
CC isolated from staphylococcus aureus. The method involves designing a  
CC series of overlapping test peptides having an amino acid sequence  
CC corresponding to (I), modifying test peptides which are identified to  
CC comprise one or more T-cell epitopes, such that they are reduced or  
CC eliminated, and repeating the T-cell eliminating modifications for (I)  
CC to produce a modified peptide or protein. Methods from the present  
CC invention can be used for the treatment, diagnosis or prophylaxis or  
CC for the preparation of a pharmaceutical composition for the treatment,  
CC diagnosis or prophylaxis of a human subject. Staphylokinase is used as a  
CC potent thrombolytic agent in patients with acute myocardial infarction.  
CC The method is useful for reducing cell based immunogenicity of non-human  
CC proteins such as streptokinase or antibodies or their fragments, from  
CC other species, for diagnostics and treatment of human disease. AAB99400  
CC to AAB99449 represent amino acid sequences used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 136 AA;  
XX

Query Match 100.0%; Score 715; DB 22; Length 136;  
Best Local Similarity 100.0%; Pred. No. 1.9e-70;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 SSSFDDGKYYKKGDASFEPTGPYLMVNTGVDSKGNELSPHYVEFPPIKPGTTLTKEKI 60  
Db 1 sssfdkgkykkgdasyfeptgpylmvntgvdsksgnellsphylvefpikpgtltlkeki 60  
XX  
QY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
Db 61 eyvewaldataykefrvvelpsakievtyydknkkkeetskspitekgfvvpdlsehi 120  
XX  
QY 121 KNPGFNLTIKVIEKK 136  
Db 121 knpgfnltikvlekk 136  
XX

RESULT 4  
AAB61909  
ID AAB61909 standard; Protein; 136 AA.  
XX  
AC AAB61909;  
XX  
DT 08-MAY-2001 (first entry)  
XX  
DE S. aureus staphylokinase G34S variant.  
XX  
KW Immunogenicity; staphylokinase; variant; stability.  
XX  
OS Staphylococcus aureus.  
XX  
FH key Location/Qualifiers  
FH Misc-difference 34  
FT /label= G34S  
FT /note= "wild-type Gly is replaced by Ser"  
XX  
PN WO200104287-A1.  
XX  
PD 18-JAN-2001.  
XX  
PF 06-JUL-2000; 2000WO-DK00371.  
XX

```
XX 07-JUL-1999; 99DK-0000988.
PR 27-AUG-1999; 99DK-0001196.
PR 02-MAR-2000; 2000DK-0000339.
PR 18-MAY-2000; 2000DK-0000804.
XX
PA (MAXY-) MAXYGEN APS.
XX
PI Halkier T, Pedersen AH, Okkels JS;
XX
DR WPI: 2001-138342/14.
XX
PT Producing polypeptides with altered immunogenicity or improved
PT stability, comprises expressing a diversified nucleotide sequence
PT population and selecting polypeptides with altered immunogenicity or
PT improved stability.
XX
PS Example 1; Page 76; 83pp; English.
XX
CC The invention relates to a method of altering immunogenicity and/or
CC increasing stability of a polypeptide of interest. The method comprises
CC (a) expressing a diversified population of nucleotide sequences encoding
CC a polypeptide of interest; (b) screening the polypeptides expressed for
CC function, immunogenicity and/or stability; and (c) selecting functional
CC polypeptides with altered immunogenicity and/or increased stability. The
CC method is used to improve the properties of polypeptides, in particular
CC to alter the immunogenicity and/or increase the functional in vivo half-
CC life of the polypeptide. The method uses a high throughput system that
CC makes it possible to search several orders of magnitude more polypeptides
CC than is possible by previously known approaches. This enhances the chance
CC of finding the optimal variant from the many thousands of variants that
CC may be produced. The present sequence represents S. aureus staphylokinase
CC variant sequence, used to exemplify the method of the invention.
XX
SQ Sequence 136 AA;

Query Match 100.0%; Score 715; DB 22; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.9e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSSFDKGYKKGGDDASYFEPTGPYLMVNVTVGVDKGNELLSPHYVEFPKPGTTLTKEKI 60
Db 1 sssfdkgykkggddasyfepgpylmvntvgvdskgnellspkyvefpikpgtlttkeki 60
OY 61 EYVVEWALDATAYKKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFVVPDLSEHI 120
Db 61 eyvvewaldataykefrvvelgpsakievtyydknkkkeetksfpitekgyfvvpdlsehi 120
OY 121 KNPGFNLTQVIEKK 136
Db 121 knpgfnltkvlekk 136

RESULT 5
AAR39150
ID AAR39150 standard; Protein; 137 AA.
XX
AC AAR39150;
XX
DT 03-DEC-1993 (first entry)
XX
DE Staphylokinase SAK-STAR.
XX
KW Signal sequence, plasminogen activator; thrombosis; staphylokinase;
KW SAK.
XX
OS Staphylococcus aureus strain 23.
XX
PN WO9313209-A.
XX
PD 08-JUL-1993.
XX
```

```
PF 28-DEC-1992; 92WO-EP02989.
XX
PR 30-DEC-1991; 91DE-4143279.
PR 22-JUN-1992; 92DE-4220516.
PR 01-DEC-1992; 92DE-4240801.
XX
PA (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.
XX
PI Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
XX
DR WPI: 1993-227325/28.
DR N-PSDB; AAQ44270.
XX
PT Staphylo-kinase (SAK) sequences lacking sequences for signal
PT peptide(s) - for prodn. of proteins used as plasminogen
PT activators in thrombosis treatment, and monoclonal antibodies
PT against SAK
XX
PS Claim 3; Fig 3; 99pp; German.
XX
CC DNA encoding SAK lacking the signal peptide, is expressed
CC intracellularly. This avoids the problem of fast degradation of the
CC polypeptides or destruction of the host when expressed into the
CC medium or into the periplasm respectively. High expression is
CC possible and the chemically induced overprodn. is easy to handle.
CC Also, the prods. are homogeneous.
CC SAK-polypeptide derivs. are plasminogen activators for the
CC treatment of thrombosis.
XX
SQ Sequence 137 AA;

Query Match 100.0%; Score 715; DB 14; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.9e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSSFDKGYKKGGDDASYFEPTGPYLMVNVTVGVDKGNELLSPHYVEFPKPGTTLTKEKI 60
Db 2 sssfdkgykkggddasyfepgpylmvntvgvdskgnellspkyvefpikpgtlttkeki 61
OY 61 EYVVEWALDATAYKKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFVVPDLSEHI 120
Db 62 eyvvewaldataykefrvvelgpsakievtyydknkkkeetksfpitekgyfvvpdlsehi 121
OY 121 KNPGFNLTQVIEKK 136
Db 122 knpgfnltkvlekk 137

RESULT 6
AAI15024
ID AAI15024 standard; Protein; 136 AA.
XX
AC AAI15024;
XX
DT 03-NOV-1999 (first entry)
XX
DE Staphylokinase variant SakSTAR (V132L).
XX
KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
KW SakSTAR (V132L); anti-thrombotic; fibrinolytic; cardiac; veterinary;
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;
KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
XX
OS Staphylococcus aureus.
OS Synthetic.
XX
FH key location/Qualifiers
FT Misc-difference 132
FT /note= "Wild type Val is substituted by Leu"
XX
PN WO9940198-A2.
XX
```



```
PD 12-AUG-1999.
XX
PF 04-FEB-1999; 99WO-EP00748.
XX
PR 06-FEB-1998; 98EP-0200365.
PR 04-FEB-1998; 98EP-0200323.
XX
PA (COLL/) COLLEN D J.
PA (LEUV-) LEUVEN RES & DEV VZW.
XX
PI Collen DJ;
XX
DR WPI; 1999-508504/42.
XX
PT Staphylokinase derivatives with reduced immunogenicity, used for,
PT e.g. treatment of arterial thrombosis
XX
PS Claim 7; Page -: 101pp; English.
XX
CC The present sequence is a specifically claimed Staphylokinase SakSTAR
CC variant. This variant has one aminoacid that has been substituted
CC by another aminoacid that reduces the reactivity with monoclonal
CC antibodies and absorption of SakSTAR-specific antibodies from plasma
CC of patients treated with staphylokinase. The derivatives can also be
CC substituted with cysteine modified with PEG to maintain the specific
CC activity and significantly reduce the plasma clearance. They have
CC altered immunogenicity without markedly reducing the specific activity.
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
CC potency in human plasma. The new staphylokinase derivatives are used
CC for treatment of arterial thrombosis, especially myocardial infarction.
CC The compositions can be used in human or veterinary practice.
CC Note: The present sequence is not shown in the specification, but is
CC derived from the Staphylococcus aureus wild type staphylokinase sequence
CC given in figure 1.
XX
SQ Sequence 136 AA;

Query Match 99.68; Score 712; DB 20; Length 136;
Best Local Similarity 99.3%; Pred. No. 4e-70;
Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSFDDKGGKKGDASYFEPTGPLYMNVNTGVDKGNELSPHYVEFPKPGTTLTKRKI 60
Db 1 sssfdkgkykddasyfepgpylmvntgvdskgnelisphvefpikpgtlttkeki 60

QY 61 EYVWEALDATAYKKEFRVVELDPSAKIEVTYYDKNKKKEETKSEFITEKGFVVPDLSEHI 120
Db 61 eyvewaldataykefrvvelpsakievtyydknkkkeetksfitekgfvpdlsehi 120

QY 121 KNPGFNLTQKVIKK 136
Db 121 knpgfnltkvliekk 136

RESULT 7
ID AAW44696 standard; protein; 136 AA.
XX
AC AAW44696;
XX
DT 01-MAY-1998 (first entry)
XX
DE Staphylokinase from Staphylococcus aureus.
XX
KW Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.
XX
OS Staphylococcus aureus.
XX
PN JP10028587-A.
XX
PD 03-FEB-1998.
XX
```

```
PF 19-JUL-1996; 96JP-0208991.
XX
PR 19-JUL-1996; 96JP-0208991.
XX
PA (HONS ) YAKULT HONSHA KK.
XX
DR WPI; 1998-162525/15.
XX
PT Staphylokinase mutants - useful as thrombolytic agents
XX
PS Example 1; Page 1; 9pp; Japanese.
XX
CC The present sequence represents staphylokinase (SAK). The invention
CC relates to seven specifically claimed staphylokinase mutants which are
CC useful as thrombolytic agents. The mutants are: (2) a SAK mutant (D41N)
CC in which the 14th Asp in the SAK which comprises 136 amino acids is
CC replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th Glu in SAK is
CC replaced by Gln; (4) a SAK mutant (K113A) in which the 86th Lys in SAK
CC is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th Lys in SAK
CC is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th Glu in SAK
CC is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th Asp in SAK
CC is replaced by Glu, and (8) a SAK mutant (K136A) in which the 109th Lys
CC in SAK is replaced by Ala.
XX
SQ Sequence 136 AA;
```

```
Query Match 99.48; Score 711; DB 19; Length 136;
Best Local Similarity 99.3%; Pred. No. 5.2e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 SSSFDDKGGKKGDASYFEPTGPLYMNVNTGVDKGNELSPHYVEFPKPGTTLTKRKI 60
Db 1 sssfdkgkykddasyfepgpylmvntgvdskgnelisphvefpikpgtlttkeki 60
```

```
QY 61 EYVWEALDATAYKKEFRVVELDPSAKIEVTYYDKNKKKEETKSEFITEKGFVVPDLSEHI 120
Db 61 eyvewaldataykefrvvelpsakievtyydknkkkeetksfitekgfvpdlsehi 120
```

```
QY 121 KNPGFNLTQKVIKK 136
Db 121 knpgfnltkvliekk 136
```

```
RESULT 8
ID AAY15022 standard; protein; 136 AA.
XX
AC AAY15022;
```

```
DT 03-NOV-1999 (first entry)
```

```
DE Staphylokinase variant SakSTAR (K74Q).
```

```
XX
KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
KW SakSTAR (K74Q); anti-thrombotic; fibrinolytic; cardiant; veterinary;
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;
KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
```

```
XX
OS Staphylococcus aureus.
XX
OS Synthetic.
```

```
XX
FH Key Location/Qualifiers
FT Misc-difference 74 /note= "Wild type Lys is substituted by Gln"
```

```
XX
PN WO9940198-A2.
```

```
XX
PD 12-AUG-1999.
```

```
XX
PF 04-FEB-1999; 99WO-EP00748.
```

```
XX
PR 06-FEB-1998; 98EP-0200365.
```

PR 04-FEB-1998; 98EP-0200323.  
XX  
XX (COLL/) COLLEN D J.  
PA (LEUV-) LEUVEN RES & DEV VZW.  
XX  
XX Colleen DJ;  
PI  
DR WPI; 1999-508504/42.  
XX  
PT Staphylokinase derivatives with reduced immunogenicity, used for,  
PT e.g. treatment of arterial thrombosis  
XX  
XX Claim 7; Page -: 101pp; English.  
XX  
CC The present sequence is a specifically claimed staphylokinase SakSTAR  
CC variant. This variant has one aminoacid that has been substituted  
CC by another aminoacid that reduces the reactivity with monoclonal  
CC antibodies and absorption of SakSTAR-specific antibodies from plasma  
CC of patients treated with staphylokinase. The derivatives can also be  
CC substituted with cysteine modified with PEG to maintain the specific  
CC activity and significantly reduce the plasma clearance. They have  
CC altered immunogenicity without markedly reducing the specific activity.  
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic  
CC potency in human plasma. The new staphylokinase derivatives are used  
CC for treatment of arterial thrombosis, especially myocardial infarction.  
CC The compositions can be used in human or veterinary practice.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the Staphylococcus aureus wild type staphylokinase sequence  
CC given in figure 1.  
XX  
SQ Sequence 136 AA;  
  
Query Match 99.4%; Score 711; DB 20; Length 136;  
Best Local Similarity 99.3%; Pred. No. 5.2e-70;  
Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SSSFDKGGKKGGDASYFEPTGPLYMNVNTGVDKGNELSPHYVEFPKPGTTLTKEKI 60  
Db 1 sssfdkgykkgddasyfepgplymnnvtgyvdsksgnellsphyvefpkpgtltlkeki 60  
  
QY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFVVPDLSEHI 120  
Db 61 eyvewaldataykefrvveldp sakievttydknkkkeetksfpitekgyfvvpdlsehi 120  
  
QY 121 KNPGFNLITKVIIEKK 136  
Db 121 knpgfnlittkvliekk 136  
  
RESULT 9  
AA15025  
ID AAY15025 standard; Protein; 136 AA.  
XX  
XX AAY15025;  
AC  
XX  
XX 03-NOV-1999 (first entry)  
DT  
XX  
DE Staphylokinase variant SakSTAR (V132T).  
XX  
XX Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;  
KW SakSTAR (V132T); anti-thrombotic; fibrinolytic; cardiant; veterinary;  
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;  
KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.  
XX  
XX Staphylococcus aureus.  
OS Synthetic.  
OS  
XX  
XX Key Location/Qualifiers  
FT Misc-difference 132  
FT /note- "Wild type Val is substituted by Thr"  
XX  
XX  
PN W09940198-A2.

XX  
PD 12-AUG-1999.  
XX  
XX 04-FEB-1999; 99WO-EP00748.  
PE  
XX  
XX 06-FEB-1998; 98EP-0200365.  
PR 04-FEB-1998; 98EP-0200323.  
XX  
XX (COLL/) COLLEN D J.  
PA (LEUV-) LEUVEN RES & DEV VZW.  
XX  
XX Colleen DJ;  
PI  
DR WPI; 1999-508504/42.  
XX  
XX  
PT Staphylokinase derivatives with reduced immunogenicity, used for,  
PT e.g. treatment of arterial thrombosis  
XX  
XX Claim 7; Page -: 101pp; English.  
XX  
CC The present sequence is a specifically claimed staphylokinase SakSTAR  
CC variant. This variant has one aminoacid that has been substituted  
CC by another aminoacid that reduces the reactivity with monoclonal  
CC antibodies and absorption of SakSTAR-specific antibodies from plasma  
CC of patients treated with staphylokinase. The derivatives can also be  
CC substituted with cysteine modified with PEG to maintain the specific  
CC activity and significantly reduce the plasma clearance. They have  
CC altered immunogenicity without markedly reducing the specific activity.  
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic  
CC potency in human plasma. The new staphylokinase derivatives are used  
CC for treatment of arterial thrombosis, especially myocardial infarction.  
CC The compositions can be used in human or veterinary practice.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the Staphylococcus aureus wild type staphylokinase sequence  
CC given in figure 1.  
XX  
SQ Sequence 136 AA;  
  
Query Match 99.4%; Score 711; DB 20; Length 136;  
Best Local Similarity 99.3%; Pred. No. 5.2e-70;  
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 SSSFDKGGKKGGDASYFEPTGPLYMNVNTGVDKGNELSPHYVEFPKPGTTLTKEKI 60  
Db 1 sssfdkgykkgddasyfepgplymnnvtgyvdsksgnellsphyvefpkpgtltlkeki 60  
  
QY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFVVPDLSEHI 120  
Db 61 eyvewaldataykefrvveldp sakievttydknkkkeetksfpitekgyfvvpdlsehi 120  
  
QY 121 KNPGFNLITKVIIEKK 136  
Db 121 knpgfnlittkvliekk 136  
  
RESULT 10  
AAY15016  
ID AAY15016 standard; Protein; 136 AA.  
XX  
XX AAY15016;  
AC  
XX  
XX 03-NOV-1999 (first entry)  
DT  
XX  
DE Staphylokinase variant SakSTAR (V132A).  
XX  
XX Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;  
KW SakSTAR (V132A); anti-thrombotic; fibrinolytic; cardiant; veterinary;  
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;  
KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.  
XX  
XX Staphylococcus aureus.  
OS Synthetic.  
OS

```

XX FH Key Location/Qualifiers
XX FT Misc-difference 132
XX FT /note= "Wild type Val is substituted by Ala"
XX PN W09940198-A2.
XX XX 12-AUG-1999.
XX PD
XX PF 04-FEB-1999; 99WO-EP00748.
XX PR 06-FEB-1998; 98EP-0200365.
XX PR 04-FEB-1998; 98EP-0200323.
XX PA (COLL/) COLLEN D J.
XX PA (LEUV-) LEUVEN RES & DEV VZW.
XX PI
XX PI Collen DJ;
XX DR WPI; 1999-508504/42.
XX PT
XX PT Staphylokinase derivatives with reduced immunogenicity, used for,
XX PT e.g. treatment of arterial thrombosis
XX PS Claim 7; Page -; 101pp; English.
XX CC The present sequence is a specifically claimed staphylokinase SakSTAR
XX CC variant. This variant has one aminoacid that has been substituted
XX CC by another aminoacid that reduces the reactivity with monoclonal
XX CC antibodies and absorption of SakSTAR-specific antibodies from plasma
XX CC of patients treated with staphylokinase. The derivatives can also be
XX CC substituted with cysteine modified with PEG to maintain the specific
XX CC activity and significantly reduce the plasma clearance. They have
XX CC altered immunogenicity without markedly reducing the specific activity.
XX CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
XX CC potency in human plasma. The new staphylokinase derivatives are used
XX CC for treatment of arterial thrombosis, especially myocardial infarction.
XX CC The compositions can be used in human or veterinary practice.
XX CC Note: The present sequence is not shown in the specification, but is
XX CC derived from the Staphylococcus aureus wild type staphylokinase sequence
XX CC given in figure 1.
XX SQ Sequence 136 AA;

Query Match 99.4%; Score 711; DB 20; Length 136;
Best Local Similarity 99.3%; Pred. No. 5.2e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGKYYKGGDDASYFEPTGPIYIMVNTGVDSKGNELSPHYVEFPPIKPGTTLTKEKI 60
   |||||||
Db 1 sssfdkgykkygddasyfepgpylmvntgvdsksgnelshphvefpikpgtltlkeki 60

OY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFWVPLSEHI 120
   |||||||
Db 61 eyvwealdataykefrvvelbpsakievtyydknkkkeetksfitekgfvpdlsehi 120

OY 121 KNGGFNLITKVVIEKK 136
   |||||||
Db 121 knpgfnliltkvaiekk 136

RESULT 11
AAY15021
ID AAY15021 standard; Protein; 136 AA.
XX AC AAY15021;
XX DT 03-NOV-1999 (first entry)
XX DE Staphylokinase variant SakSTAR (K35E).
XX ST Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;

```

KW	SakSTAR (K35E); anti-thrombotic; fibrinolytic; cardiant; veterinary;
KW	Staphylococcus aureus; myocardial infarction; arterial thrombosis;
KW	monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
XX	
OS	Staphylococcus aureus.
OS	Synthetic.
FH	Key Location/Qualifiers
FT	Misc-difference 35 /note= "Wild type Lys is substituted by Glu"
PN	WO940198-A2.
XX	
PD	12-AUG-1999.
XX	
PF	04-FEB-1999; 99WO-EP00748.
XX	
PR	06-FEB-1998; 98EP-0200365.
PR	04-FEB-1998; 98EP-0200323.
XX	
PA	(COLL/) COLLEN D J.
PA	(LEUV-) LEUVEN RES & DEV VZW.
XX	
PI	Collen DJ;
XX	
DR	WPI; 1999-508504/42.
XX	
PT	Staphylokinase derivatives with reduced immunogenicity, used for,
PT	e.g. treatment of arterial thrombosis
XX	
PS	Claim 7; Page -: 101pp; English.
XX	
CC	The present sequence is a specifically claimed Staphylokinase SakSTAR
CC	variant. This variant has one aminoacid that has been substituted
CC	by another aminoacid that reduces the reactivity with monoclonal
CC	antibodies and absorption of SakSTAR-specific antibodies from plasma
CC	of patients treated with staphylokinase. The derivatives can also be
CC	substituted with cysteine modified with PEG to maintain the specific
CC	activity and significantly reduce the plasma clearance. They have
CC	altered immunogenicity without markedly reducing the specific activity.
CC	This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
CC	potency in human plasma. The new staphylokinase derivatives are used
CC	for treatment of arterial thrombosis, especially myocardial infarction.
CC	The compositions can be used in human or veterinary practice.
CC	Note: The present sequence is not shown in the specification, but is
CC	derived from the Staphylococcus aureus wild type staphylokinase sequence
CC	given in figure 1.
XX	
SQ	Sequence 136 AA;
Query Match	99.4%; Score 711; DB 20; Length 136;
Best Local Similarity	99.3%; Pred. No. 5.2e-70;
Matches 135; Conservative	1; Mismatches 0; Indels 0; Gaps 0;
OY	1 SSSFDKGKKYGDDASYFETGPLYLMVNTGVDSKGNELSPHYVEFPPIKPGTTLTKKI 60 
Db	1 sssfdkgkykkgddasyfptgplylmvntgvdsgegnellsphvefpikpgtlttkki 60 
OY	61 EYYEWALDATAYKEFRVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFVPDLSEHI 120 
Db	61 eyyewaldataykefrveldpsakievtydydknkkkeetksfpitekgyfvpldsehi 120 
OY	121 KNPGFNLTIKVIEKK 136 
Db	121 knpgfnliltkviekk 136 
RESULT 12	
AAB61908	
ID AAB61908 standard; Protein; 136 AA.	
XX	
AC AAB61908;	

```
XX 08-MAY-2001 (first entry)
DT
XX S. aureus staphylokinase mature protein sequence.
DE
XX Immunogenicity; staphylokinase; variant; stability.
KW
XX Staphylococcus aureus.
OS
XX WO200104287-A1.
PN
XX 18-JAN-2001.
PD
XX 06-JUL-2000; 2000WO-DK00371.
PF
XX 07-JUL-1999; 99DK-0000988.
PR 27-AUG-1999; 99DK-0001196.
PR 02-MAR-2000; 2000DK-0000339.
PR 18-MAY-2000; 2000DK-0000804.
XX
XX (MAXY-) MAXYGEN APS.
PA
XX
XX Halkier T, Pedersen AH, Okkels JS;
PI
XX WPI; 2001-138342/14.
DR
XX N-PSDB; AAC85112.
DR
XX
XX Producing polypeptides with altered immunogenicity or improved
PT stability, comprises expressing a diversified nucleotide sequence
PT population and selecting polypeptides with altered immunogenicity or
PT improved stability -
XX
XX Example 1; Page 75; 83pp; English.
PS
XX The invention relates to a method of altering immunogenicity and/or
CC increasing stability of a polypeptide of interest. The method comprises
CC (a) expressing a diversified population of nucleotide sequences encoding
CC a polypeptide of interest; (b) screening the polypeptides expressed for
CC function, immunogenicity and/or stability; and (c) selecting functional
CC polypeptides with altered immunogenicity and/or increased stability. The
CC method is used to improve the properties of polypeptides, in particular
CC to alter the immunogenicity and/or increase the functional in vivo half-
CC life of the polypeptide. The method uses a high throughput system that
CC makes it possible to search several orders of magnitude more polypeptides
CC than is possible by previously known approaches. This enhances the chance
CC of finding the optimal variant from the many thousands of variants that
CC may be produced. The present sequence represents S. aureus staphylokinase
CC mature protein sequence, used to exemplify the method of the invention.
XX
XX Sequence 136 AA;
SQ
Query Match 99.4%; Score 711; DB 22; Length 136;
Best Local Similarity 99.3%; Pred. No. 5.2e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 SSSFDKGGKKKGDASFEPTGPLYMNVNTGVDSKGNELLSPHYVEFPPIKPGTTLTKEKI 60
DB 1 sssfdkgkkgddasyfeptgplymwnvtgvdgkgnellsphyvefpikpgtltlkeki 60
OY 61 EYVVEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFVVPDLSEHI 120
DB 61 eyvewaldataykefrveldpsakilevtydknkkkeetksfpitekgfvvpdlsehi 120
OY 121 KNPGFNLTITKVIEKK 136
DB 121 knpgfnlittkviekk 136
RESULT 13
AAR39149
ID AAR39149 standard; Protein; 137 AA.
XX
```

```
AC AAR39149;
XX
XX 03-DEC-1993 (first entry)
DT
XX Staphylokinase SAK-CphC.
DE
XX Signal sequence, plasminogen activator; thrombosis; staphylokinase;
KW SAK.
XX
XX Staphylococcus aureus phage phiC.
OS
XX WO9313209-A.
PN
XX 08-JUL-1993.
PD
XX 28-DEC-1992; 92WO-EP02989.
PF
XX 30-DEC-1991; 91DE-4143279.
PR 22-JUN-1992; 92DE-4220516.
PR 01-DEC-1992; 92DE-4240801.
XX
XX (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.
PA
XX
XX Albrecht S, Behnke D, Guehrs K, Hartmann M, Schloott B;
PI
XX WPI; 1993-227325/28.
DR
XX N-PSDB; AAQ44269.
DR
XX
XX Staphylo-kinase (SAK) sequences lacking sequences for signal
PT peptide(s) - for prodn. of proteins used as plasminogen
PT activators in thrombosis treatment, and monoclonal antibodies
PT against SAK
XX
XX Claim 3; Fig 2; 99pp; German.
PS
XX
XX DNA encoding SAK lacking the signal peptide, is expressed
CC intracellularly. This avoids the problem of fast degradation of the
CC polypeptides or destruction of the host when expressed into the
CC medium or into the periplasm respectively. High expression is
CC possible and the chemically induced overprodn. is easy to handle.
CC Also, the prods. are homogeneous.
CC SAK-polypeptide derivs. are plasminogen activators for the
CC treatment of thrombosis.
XX
XX Sequence 137 AA;
SQ
Query Match 99.4%; Score 711; DB 14; Length 137;
Best Local Similarity 99.3%; Pred. No. 5.3e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 SSSFDKGGKKKGDASFEPTGPLYMNVNTGVDSKGNELLSPHYVEFPPIKPGTTLTKEKI 60
DB 2 sssfdkgkkgddasyfeptgplymwnvtgvdgkgnellsphyvefpikpgtltlkeki 61
OY 61 EYVVEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFVVPDLSEHI 120
DB 62 eyvewaldataykefrveldpsakilevtydknkkkeetksfpitekgfvvpdlsehi 121
OY 121 KNPGFNLTITKVIEKK 136
DB 122 knpgfnlittkviekk 137
RESULT 14
AAR12137
ID AAR12137 standard; Protein; 163 AA.
XX
XX AAR12137;
AC
XX 05-AUG-1991 (first entry)
DT
XX S.aureus Staphylokinase with signal peptide.
XX
```



```
XX SAK: protein production.
KM Synthetic.
XX
OS
XX Key Location/Qualifiers
FH Peptide 1..27
FT Protein /label= signal peptide
FT 28..163
FT /label= staphylokinase
XX
XX JP03098595-A.
XX
XX 24-APR-1991.
XX
XX 11-SEP-1989; 89JP-0234874.
XX
XX 11-SEP-1989; 89JP-0234874.
XX
XX (TAIS ) TAISHO PHARMACEUT KK.
XX
XX WPI; 1991-167039/23.
XX
XX N-PSDB; AAQ11813.
XX
XX Prepn. of peptide(s) - by construction of expression
XX vector, transformation of E.coli etc., culturing to secrete
XX peptide(s) and collecting peptide(s)
XX
XX Example; Fig 3; 15pp; Japanese.
XX
XX The first 79 residues of this sequence form part of a fusion
XX protein with somatomedin C. A synthetic construct encoding the
XX SAK-SMC fusion is used to transform E.coli or Bacillus subtilis.
XX The SAK signal peptide directs extracellular secretion of the
XX somatomedin C.
XX See also AAQ11814.
XX
XX Sequence 163 AA;
SQ
```

```
Query Match 99.4%; Score 711; DB 12; Length 163;
Best Local Similarity 99.3%; Pred. No. 6.7e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 SSSFDKGYKKKGDASFEPTGPLYMNVNTGVDKGNELSPHYVEFPKPGTTLTKKI 60
   |||||||
Db 28 sssfdkgykkkgddasyfepptgplymnnvtgvdgkgnellsphyvefpikpgtlttkki 87

QY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120
   |||||||
Db 88 eyvewaldataykefrvvelpsakievtydydknkkkeetksfptekgfvpdlsehi 147

QY 121 KNPGFNLTITKVIEKK 136
   |||||||
Db 148 knpgfnltitkviekk 163
```

```
RESULT 15
AAR28844
ID AAR28844 standard; Protein; 163 AA.
XX
XX AAR28844;
AC
XX 13-JAN-1993 (first entry)
DT
XX Staphylokinase (SAK).
DE
XX Staphylokinase; thrombosis; fibrinogen; plasminogen.
KM
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Peptide 1..27
```

```
FT Protein /label= sig_peptide
FT 28..163
FT /label= mat_protein
XX
XX WO9211356-A.
XX
XX 09-JUL-1992.
PD
XX 17-DEC-1991; 91WO-JP01722.
XX
XX 17-DEC-1990; 90JP-0411063.
XX
XX (HONS ) YAKULT HONSHA KK.
XX
XX Hashimoto S, Matsumoto T, Matsuo O, Onoue M, Sakai M;
XX Sako T, Sansawa H, Shimura K, Shishido Y, Watanabe T;
XX Yokokura T;
XX WPI; 1992-250079/30.
XX
XX Thrombolytic peptide SAK-11 derived from staphylokinase - useful
XX as a plasminogen activator for treating thrombosis
XX
XX Disclosure; Page 16; 26pp; Japanese.
XX
XX SAK-11 (AAR25468) is derived from its precursor, staphylokinase (SAK)
XX - represented in AAR28844, by tryptic cleavage of its 10 N-terminal
XX amino acid residues. SAK-11 is useful as the active agent in an
XX injectable treatment for thrombosis. Thrombolytic activity is
XX demonstrated using a rabbit jugular vein thrombosis model. A clear
XX decrease in fibrinogen is observed upon application of 0.9 mg SAK-11/
XX kg. Plasminogen activation is superior to that of SAK (no specific
XX data given). Furthermore, SAK-11 has very low antigenicity as shown
XX in a PCA study, against streptokinase and egg albumin controls, in
XX BALB/c mice.
XX
XX Sequence 163 AA;
SQ
```

```
Query Match 99.4%; Score 711; DB 13; Length 163;
Best Local Similarity 99.3%; Pred. No. 6.7e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 SSSFDKGYKKKGDASFEPTGPLYMNVNTGVDKGNELSPHYVEFPKPGTTLTKKI 60
   |||||||
Db 28 sssfdkgykkkgddasyfepptgplymnnvtgvdgkgnellsphyvefpikpgtlttkki 87

QY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120
   |||||||
Db 88 eyvewaldataykefrvvelpsakievtydydknkkkeetksfptekgfvpdlsehi 147

QY 121 KNPGFNLTITKVIEKK 136
   |||||||
Db 148 knpgfnltitkviekk 163
```

```
RESULT 16
AAV15013
ID AAV15013 standard; Protein; 136 AA.
XX
XX AAV15013;
AC
XX 03-NOV-1999 (first entry)
DT
XX Staphylokinase variant SakSTAR (T90A).
DE
XX Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
XX SakSTAR (T90A); anti-thrombotic; fibrinolytic; cardiant; veterinary;
XX Staphylococcus aureus; myocardial infarction; arterial thrombosis;
XX monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
OS
XX Staphylococcus aureus.
OS Synthetic.
```

```
XX Key Location/Qualifiers
FH Misc-difference 90
FT /note= "Wild type Thr is substituted by Ala"
XX
XX WO9940198-A2.
XX
XX 12-AUG-1999.
XX
XX 04-FEB-1999; 99WO-EP00748.
XX
XX 06-FEB-1998; 98EP-0200365.
XX
XX 04-FEB-1998; 98EP-0200323.
XX
XX (COLL/) COLLEN D J.
XX (LEUV-) LEUVEN RES & DEV VZW.
XX
XX Collen DJ;
XX
XX WPI; 1999-508504/42.
XX
XX Staphylokinase derivatives with reduced immunogenicity, used for,
XX e.g. treatment of arterial thrombosis
XX
XX Claim 7; Page -; 101pp; English.
XX
XX The present sequence is a specifically claimed Staphylokinase SakSTAR
XX variant. This variant has one aminoacid that has been substituted
XX by another aminoacid that reduces the reactivity with monoclonal
XX antibodies and absorption of SakSTAR-specific antibodies from plasma
XX of patients treated with staphylokinase. The derivatives can also be
XX substituted with cysteine modified with PEG to maintain the specific
XX activity and significantly reduce the plasma clearance. They have
XX altered immunogenicity without markedly reducing the specific activity.
XX This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
XX potency in human plasma. The new staphylokinase derivatives are used
XX for treatment of arterial thrombosis, especially myocardial infarction.
XX The compositions can be used in human or veterinary practice.
XX Note: The present sequence is not shown in the specification, but is
XX derived from the Staphylococcus aureus wild type staphylokinase sequence
XX given in figure 1.
XX
XX Sequence 136 AA;

Query Match 99.3%; Score 710; DB 20; Length 136;
Best Local Similarity 99.3%; Pred. No. 6.7e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGKKKGDASYEPTGPLYMNVNVTGVDKGNELLSPHYVEFPPIKPGTTLTKERI 60
Db 1 sssfdkgkkykkgddasyfeptgplymwnvcgvdskgnellspkyvefpikpgtltlkeki 60

OY 61 EYVWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFPVPLDSEHI 120
Db 61 eyvwaldataykefrvvelldpsakilevtydydknkkkeetksfptekgfvpdlsehi 120

OY 121 KNPGFNLTITKVIEKK 136
Db 121 knpgfnlittkviekk 136

RESULT 17
AAW03101 standard; protein; 136 AA.
XX
XX AAW03101;
XX
XX 19-FEB-1997 (first entry)
XX
XX Staphylokinase derivative M18 with altered charge cluster 18.
XX
XX Staphylokinase; mutant; mutein; variant; immunogenicity; decrease;
```

```
KW derivative; SakSTAR; arterial thrombosis; thrombolytic agent.
XX
XX Staphylococcus aureus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Region 130
FT /label= altered_M18_cluster
FT /note= "wild-type cluster 18 sequence lys
FT has been replaced by Ala to reduce
FT immunogenicity"
XX
XX WO9621016-A2.
XX
XX 11-JUL-1996.
XX
XX 03-JAN-1996; 96WO-EP00081.
XX
XX 17-NOV-1995; 95JP-0299781.
XX
XX 06-JAN-1995; 95EP-0200023.
XX
XX 11-JAN-1995; 95US-0371505.
XX
XX 09-JUN-1995; 95EP-0201531.
XX
XX 06-JUL-1995; 95US-0499092.
XX
XX (COLL/) COLLEN D.
XX (LEUV-) LEUVEN RES & DEV VZW.
XX
XX Collen D;
XX
XX WPI; 1996-333991/33.
XX
XX New staphylokinase derivs. having reduced immunogenicity - useful
XX for treating arterial thrombosis
XX
XX Example 2; Page -; 58pp; English.
XX
XX Staphylokinase derivatives showing a reduced immunogenicity as
XX compared to wild-type staphylokinase are claimed. The derivatives
XX are useful as thrombolytic agents to treat arterial thrombosis and
XX are pref. produced by eliminating at least one of the epitopes
XX indicated in the features table. The epitopes are destroyed by
XX replacing one or more amino acid residues in a charge cluster by an
XX Ala residue. Mutations are introduced using site-directed
XX mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus
XX lysogenic strain SakSTAR. The present sequence is a specific
XX example of a staphylokinase derivative according to the invention;
XX the sequence does not appear in the specification but can be generated
XX from the wild-type staphylokinase sequence using the description given
XX in Table 1 of the disclosure.
XX
XX Sequence 136 AA;

Query Match 99.2%; Score 709; DB 17; Length 136;
Best Local Similarity 99.3%; Pred. No. 8.6e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGKKKGDASYEPTGPLYMNVNVTGVDKGNELLSPHYVEFPPIKPGTTLTKERI 60
Db 1 sssfdkgkkykkgddasyfeptgplymwnvcgvdskgnellspkyvefpikpgtltlkeki 60

OY 61 EYVWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFPVPLDSEHI 120
Db 61 eyvwaldataykefrvvelldpsakilevtydydknkkkeetksfptekgfvpdlsehi 120

OY 121 KNPGFNLTITKVIEKK 136
Db 121 knpgfnlittaviekk 136

RESULT 18
AAI5023 standard; Protein; 136 AA.
ID AAI5023
```

```

XX  AAY15023;
AC
XX  03-NOV-1999 (first entry)
DT
XX
XX  Staphylokinase variant SakSTAR (K130T).
DE
XX
XX  Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
KM  SakSTAR (K130T); anti-thrombotic; fibrinolytic; cardiant; veterinary;
KM  Staphylococcus aureus; myocardial infarction; arterial thrombosis;
KM  monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
XX
OS  Staphylococcus aureus.
OS  Synthetic.
XX
FH  Key Location/Qualifiers
FT  Misc-difference 130
FT  /note= "Wild type Lys is substituted by Thr"
XX
XX  W09940198-A2.
XX
XX  12-AUG-1999.
XX
XX  04-FEB-1999; 99WO-EP00748.
XX
XX  06-FEB-1998; 98EP-0200365.
XX  04-FEB-1998; 98EP-0200323.
XX
XX  (COLL/) COLLEN D J.
PA  (LEUV-) LEUVEN RES & DEV VZW.
XX
XX  Collen DJ;
PI
XX
XX  WPI; 1999-508504/42.
XX
XX  Staphylokinase derivatives with reduced immunogenicity, used for,
PT  e.g. treatment of arterial thrombosis
XX
XX  Claim 7; Page -; 101pp; English.
XX
XX  The present sequence is a specifically claimed Staphylokinase SakSTAR
CC  variant. This variant has one aminoacid that has been substituted
CC  by another aminoacid that reduces the reactivity with monoclonal
CC  antibodies and absorption of SakSTAR-specific antibodies from plasma
CC  of patients treated with staphylokinase. The derivatives can also be
CC  substituted with cysteine modified with PEG to maintain the specific
CC  activity and significantly reduce the plasma clearance. They have
CC  altered immunogenicity without markedly reducing the specific activity.
CC  This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
CC  potency in human plasma. The new staphylokinase derivatives are used
CC  for treatment of arterial thrombosis, especially myocardial infarction.
CC  The compositions can be used in human or veterinary practice.
CC  Note: The present sequence is not shown in the specification, but is
CC  derived from the Staphylococcus aureus wild type staphylokinase sequence
CC  given in figure 1.
XX
XX  Sequence 136 AA;
SQ

Query Match 99.2%; Score 709; DB 20; Length 136;
Best Local Similarity 99.3%; Pred. No. 8.6e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGGKYYKGGDDASYFEPTGPIYMNVTGVDKSGNELLSPHYVEFPPIKPGTTLTKKEKI 60
    |||||||
Dd 1 sssfdkgkyykgddasyfeptgpylmvntvgvdsksgnellsphyvefpikpgtlttkeki 60
    |||||||

OY 61 EYYEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFWVPDLSEHI 120
    |||||||
Dd 61 eyyewaldataykefrvveldpakievttydydknkkkeetksfptekgfvvpdlsehi 120
    |||||||

OY 121 KNPGFNLTIRKVIKK 136
    |||||||

```

```

Db          121 knpgfnlittvlekk 136

RESULT 19
ID AAY15015
AC AAY15015;
DT 03-NOV-1999 (first entry)
DE Staphylokinase variant SakSTAR (K130A).
KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;
KW SakSTAR (K130A); anti-thrombotic; fibrinolytic; cardiant; veterinary;
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;
KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.
XX
OS Staphylococcus aureus.
OS Synthetic.
FH key Location/Qualifiers
FT Misc-difference 130
FT /note= "Wild type Lys is substituted by Ala"
XX
PN W09940198-A2.
XX
PD 12-AUG-1999.
XX
PF 04-FEB-1999; 99WO-EP00748.
XX
PR 06-FEB-1998; 98EP-0200365.
PR 04-FEB-1998; 98EP-0200323.
XX
PA (COLL/) COLLEN D J.
PA (LEUV-) LEUVEN RES & DEV VZW.
XX
PI Collen DJ;
XX
DR WPI; 1999-508504/42.
XX
PT Staphylokinase derivatives with reduced immunogenicity, used for,
PT e.g. treatment of arterial thrombosis
XX
PS Claim 7; Page -: 101pp; English.
XX
CC The present sequence is a specifically claimed Staphylokinase SakSTAR
CC variant. This variant has one aminoacid that has been substituted
CC by another aminoacid that reduces the reactivity with monoclonal
CC antibodies and absorption of SakSTAR-specific antibodies from plasma
CC of patients treated with staphylokinase. The derivatives can also be
CC substituted with cysteine modified with PEG to maintain the specific
CC activity and significantly reduce the plasma clearance. They have
CC altered immunogenicity without markedly reducing the specific activity.
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
CC potency in human plasma. The new staphylokinase derivatives are used
CC for treatment of arterial thrombosis, especially myocardial infarction.
CC The compositions can be used in human or veterinary practice.
CC Note: The present sequence is not shown in the specification, but is
CC derived from the staphylococcus aureus wild type staphylokinase sequence
CC given in figure 1.
XX
SQ Sequence 136 AA;

Query Match 99.2%; Score 709; DB 20; Length 136;
Best Local Similarity 99.3%; Pred. No. 8.6e-70;
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSSPDKGKYKKGGDDASYFEPTGPRYAMVNTGVDSKGNELSPHYVEFPPIKPGTTLTKERI 60
   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   1 sssfdkgykyykddasyfeptgpylmwnvtgvdskgnellspkhyvefpikpgtlttkeki 60

```





CC	given in figure 1.
XX	
SQ	Sequence 136 AA;

**SQ Sequence 136 AA;**

Query Match	99.28;	Score 709;	DB 20;	Length 136;
Best Local Similarity	99.38;	Pred. No. 8.6e-70;		
Matches 135; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

	QY	1	SSSFDKGKYKKKGDASYFEPTGPYLMAVNTGVDSKGNELSPHYVEEPPIKPCTTLTREKI	60
	Db	1	sssfdkgkykkkgdasyfepptgpymavntgvdskgnelshpyvefpikpgtltrekeli	60
	QY	61	EYYEWALDAPAYKEFRVELDPASAKIEVTYYYDNKKKKKETKSPITEKGFWVPDLSEHI	120
	Db	61	eyyewaldapaykefrveldpasakievtydydnkknkkeketskspitekgfwvpdlsehi	120
	QY	121	KNPGFNLTITKVIEKK	136
	Db	121	knpqfnlittkviekk	136

## RESULT 22

AA15006  
ID AA15006 standard; Protein; 136 AA.

AC AAY15006;

DT 03-NOV-1999 (first entry)

DE Staphylokinase variant SakSTAR (E80A).

KW Staphylokinase; variant; immunogenic; specificity; SakSTAR (E80A);  
 KW derivative; anti-thrombotic; fibrinolytic; cardiant; veterinary;  
 KW *Staphylococcus aureus*; myocardial infarction; arterial thrombosis; PEG;  
 KW monoclonal antibody; cysteine; polyethylene glycol; plasma clearance.  
 XX  
 OS *Staphylococcus aureus*.  
 OS Synthetic.

FH	Key	Location/Qualifiers
FT	Misc-difference	80
FT		/note- "Wild type Glu is substituted by Ala"

PN W09940198-A2

PD 12-AUG-1999

PF 04-FEB-1999; 99WO-EP00748.

PR 06-FEB-1998; 98EP-0200365.

PR 04-FEB-1998; 98EP-0200323.

PA (COLL/) COLLEN D J.  
PA (LEV-) LEUVEN RES & DEV VZW.

PI collen DJ;

DR WPI; 1999-508504/42.

PT Staphylokinase derivatives with reduced immunogenicity, used for,  
PT e.g. treatment of arterial thrombosis

PS Claim 7; Page -; 101pp; English

CC The present sequence is a specifically claimed Staphylokinase SakSTAR  
CC variant. This variant has one aminoacid that has been substituted  
CC by another aminoacid that reduces the reactivity with monoclonal  
CC antibodies and absorption of SakSTAR-specific antibodies from plasma  
CC of patients treated with staphylokinase. The derivatives can also be  
CC substituted with cysteine modified with PEG to maintain the specific  
CC activity and significantly reduce the plasma clearance. They have  
CC altered immunogenicity without markedly reducing the specific activity

CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic  
CC potency in human plasma. The new staphylokinase derivatives are used  
CC for treatment of arterial thrombosis, especially myocardial infarction.  
CC The compositions can be used in human or veterinary practice.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the *Staphylococcus aureus* wild type staphylokinase sequence  
CC given in figure 1.

**SQ Sequence 136 AA;**

Query Match	99.28;	Score 709;	DB 20;	Length 136;
Best Local Similarity	99.38;	Pred. No. 8.6e-70;		
Matches 135; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

[illegible]

## RESULT 23

ID AAW44690 standard; protein; 136 AA.

AC AAW44690;

DT 11-MAY-1998 (first entry)

DE Staphylokinase mutant (E85Q), useful as a thrombolytic agent.

**KW** Staphylokinase mutant; thrombolytic agent; *Staphylococcus aureus*.

05 staphylococcus aureus.  
05 synthetic.

FT	Key	58	/note= "Wild-type Glu has been replaced by Gln"
FT	Modified-site		
FT			

PN JP10028587-A.

PD 03-FEB-1998.

PF 19-JUL-1996; 96JP-0108991.

PR 19-JUL-1996; 96JP-0208991.

PA (HONS ) YAKULT HONSHA KK.

DR WPI; 1998-162525/15.

PT	Staphylokinase mutants - useful as thrombolytic agents
XX	
PS	Claim 1; Page 6; 9pp; Japanese.

CC The present sequence represents a specifically claimed staphylokinase  
CC (SAK) mutant (E85Q). The invention relates to seven staphylokinase  
CC mutants which are useful as thrombolytic agents. The mutants are: (2) a  
CC SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136  
CC amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th  
CC Glu in SAK is replaced by Gln; (4) a SAK mutant (K113A) in which the 86th  
CC Lys in SAK is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th  
CC Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th  
CC Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th  
CC Asp in SAK is replaced by Glu; and (8) a SAK mutant (K136A) in which the



CC given in figure 1.  
XX Sequence 136 AA;

Query Match 99.0%; Score 708; DB 20; Length 136;  
Best Local Similarity 99.3%; Pred. No. 1.1e-69;  
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGYKKGDASVFEPTGYLMVNTGVDSKGNELSPHYVEFPKPGTTLTKEKI 60  
|||||  
Db 1 sssfdkgykkgddasyfepitgpylmvntgvdsksgnellsphvefpikpgtlttkeki 60  
OY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
|||||  
Db 61 eyvewaldataykefrvvelldpsakievtyydknkkkeetksfptekgfivvpdlsehi 120  
OY 121 KNPGFNLITKVVIEKK 136  
|||||  
Db 121 knpgfnliltkvnlekk 136

## RESULT 26

AAV15027  
ID AAV15027 standard; Protein; 136 AA.

AC AAV15027;

DT 03-NOV-1999 (first entry)

DE Staphylokinase variant SakSTAR (V132R).

XX Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;  
KW SakSTAR (V132R); anti-thrombotic; fibrinolytic; cardiant; veterinary;  
KM Staphylococcus aureus; myocardial infarction; arterial thrombosis;  
KM monoclonal antibody; polyethylene glycol; PEG; plasma clearance.  
XX Staphylococcus aureus.  
OS Synthetic.

FH Key Location/Qualifiers  
FT Misc-difference 132 /note= "Wild type Val is substituted by Arg"  
FT

PN WO9940198-A2.

PD 12-AUG-1999.

PF 04-FEB-1999; 99WO-EP00748.

PR 06-FEB-1998; 98EP-0200365.

PR 04-FEB-1998; 98EP-0200323.

PA (COLL/) COLLEN D J.  
PA (LEUV-) LEUVEN RES & DEV VZW.

PI Collen DJ;

DR WPI; 1999-508504/42.

PT Staphylokinase derivatives with reduced immunogenicity, used for,  
PT e.g. treatment of arterial thrombosis

PS Claim 7; Page -, 101pp; English.

CC The present sequence is a specifically claimed Staphylokinase SakSTAR  
CC variant. This variant has one aminoacid that has been substituted  
CC by another aminoacid that reduces the reactivity with monoclonal  
CC antibodies and absorption of SakSTAR-specific antibodies from plasma  
CC of patients treated with staphylokinase. The derivatives can also be  
CC substituted with cysteine modified with PEG to maintain the specific  
CC activity and significantly reduce the plasma clearance. They have  
CC altered immunogenicity without markedly reducing the specific activity.

CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic  
CC potency in human plasma. The new staphylokinase derivatives are used  
CC for treatment of arterial thrombosis, especially myocardial infarction.  
CC The compositions can be used in human or veterinary practice.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the Staphylococcus aureus wild type staphylokinase sequence  
CC given in figure 1.

XX Sequence 136 AA;

Query Match 99.0%; Score 708; DB 20; Length 136;  
Best Local Similarity 99.3%; Pred. No. 1.1e-69;  
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGYKKGDASVFEPTGYLMVNTGVDSKGNELSPHYVEFPKPGTTLTKEKI 60  
|||||  
Db 1 sssfdkgykkgddasyfepitgpylmvntgvdsksgnellsphvefpikpgtlttkeki 60  
OY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
|||||  
Db 61 eyvewaldataykefrvvelldpsakievtyydknkkkeetksfptekgfivvpdlsehi 120  
OY 121 KNPGFNLITKVVIEKK 136  
|||||  
Db 121 knpgfnliltkvnlekk 136

## RESULT 27

AAW44694  
ID AAW44694 standard; protein; 136 AA.

AC AAW44694;

DT 11-MAY-1998 (first entry)

DE Staphylokinase mutant (D96E), useful as a thrombolytic agent.

XX Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.

OS Synthetic.

FH Key Location/Qualifiers  
FT Modified-site 69 /note= "Wild-type Asp has been replaced by Glu"  
FT

PN JP10028587-A.

PD 03-FEB-1998.

PF 19-JUL-1996; 96JP-0108991.

PR 19-JUL-1996; 96JP-0208991.

PA (HONS ) YAKULT HONSHA KK.

DR WPI; 1998-162525/15.

PT Staphylokinase mutants - useful as thrombolytic agents

PS Claim 1; Page 8; 9pp; Japanese.

CC The present sequence represents a specifically claimed staphylokinase  
CC (SAK) mutant (D96E). The invention relates to seven staphylokinase  
CC mutants which are useful as thrombolytic agents. The mutants are: (2) a  
CC SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136  
CC amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th  
CC Glu in SAK is replaced by Gln; (4) a SAK mutant (K113A) in which the 86th  
CC Lys in SAK is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th  
CC Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th  
CC Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th  
CC Asp in SAK is replaced by Glu, and (8) a SAK mutant (K136A) in which the





CC by another aminoacid that reduces the reactivity with monoclonal  
CC antibodies and absorption of SakSTAR-specific antibodies from plasma  
CC of patients treated with staphylokinase. The derivatives can also be  
CC substituted with cysteine modified with PEG to maintain the specific  
CC activity and significantly reduce the plasma clearance. They have  
CC altered immunogenicity without markedly reducing the specific activity.  
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic  
CC potency in human plasma. The new staphylokinase derivatives are used  
CC for treatment of arterial thrombosis, especially myocardial infarction.  
CC The compositions can be used in human or veterinary practice.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the Staphylococcus aureus wild type staphylokinase sequence  
CC given in figure 1.  
XX  
SQ Sequence 136 AA;

Query Match 98.9%; Score 707; DB 20; Length 136;  
Best Local Similarity 99.3%; Pred. No. 1.4e-69;  
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGYKKGGDASYFEPTGPLYMWNVTGVDKSGNELSPHYVEPIKPGTTLTKEKI 60  
|||||  
Db 1 sssfdkgykkgddasyfepgtplymwnvtgvdsknelssphyvefpikpgtltlkeki 60  
OY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSFPITEKGFVVPDLSEHI 120  
|||||  
Db 61 eyvewaldataykefrvvelpsakievtydknkkkeetksfpitekgyfvvpdlsehl 120  
OY 121 KNPGFNLTAKVIEKK 136  
|||||  
Db 121 knpgfnltakviek 136

RESULT 30

AAI15019  
ID AAI15019 standard; Protein; 136 AA.

XX  
AC AAI15019;

DT 03-NOV-1999 (first entry)

DE Staphylokinase variant SakSTAR (H43R).

KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;

KW SakSTAR (H43R); anti-thrombotic; fibrinolytic; cardiant; veterinary;

KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;

KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.

XX  
OS Staphylococcus aureus.

OS Synthetic.

FH Key Location/Qualifiers

FT Misc-difference 43

FT /note= "Wild type His is substituted by Arg"

XX  
PN WO9940198-A2.

PD 12-AUG-1999.

PF 04-FEB-1999; 99WO-EP00748.

PR 06-FEB-1998; 98EP-0200365.

PR 04-FEB-1998; 98EP-0200323.

XX  
PA (COLL/) COLLEN D J.

PA (LEUV-) LEUVEN RES & DEV VZW.

XX  
PI Collen DJ;

XX  
DR WPI; 1999-508504/42.

XX  
PT Staphylokinase derivatives with reduced immunogenicity, used for,

PT e.g. treatment of arterial thrombosis  
XX  
XX Claim 7; Page -; 101pp; English.

CC The present sequence is a specifically claimed staphylokinase SakSTAR  
CC variant. This variant has one aminoacid that has been substituted  
CC by another aminoacid that reduces the reactivity with monoclonal  
CC antibodies and absorption of SakSTAR-specific antibodies from plasma  
CC of patients treated with staphylokinase. The derivatives can also be  
CC substituted with cysteine modified with PEG to maintain the specific  
CC activity and significantly reduce the plasma clearance. They have  
CC altered immunogenicity without markedly reducing the specific activity.  
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic  
CC potency in human plasma. The new staphylokinase derivatives are used  
CC for treatment of arterial thrombosis, especially myocardial infarction.  
CC The compositions can be used in human or veterinary practice.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the staphylococcus aureus wild type staphylokinase sequence  
CC given in figure 1.  
XX  
SQ Sequence 136 AA;

Query Match 98.9%; Score 707; DB 20; Length 136;  
Best Local Similarity 99.3%; Pred. No. 1.4e-69;  
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGYKKGGDASYFEPTGPLYMWNVTGVDKSGNELSPHYVEPIKPGTTLTKEKI 60  
|||||  
Db 1 sssfdkgykkgddasyfepgtplymwnvtgvdsknelssphyvefpikpgtltlkeki 60  
OY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSFPITEKGFVVPDLSEHI 120  
|||||  
Db 61 eyvewaldataykefrvvelpsakievtydknkkkeetksfpitekgyfvvpdlsehl 120  
OY 121 KNPGFNLTAKVIEKK 136  
|||||  
Db 121 knpgfnltakviek 136

RESULT 31

AAI15011  
ID AAI15011 standard; Protein; 136 AA.

XX  
AC AAI15011;

DT 03-NOV-1999 (first entry)

DE Staphylokinase variant SakSTAR (D82A).

KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;

KW SakSTAR (D82A); anti-thrombotic; fibrinolytic; cardiant; veterinary;

KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;

KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.

XX  
OS Staphylococcus aureus.

OS Synthetic.

FH Key Location/Qualifiers

FT Misc-difference 82

FT /note= "Wild type Asp is substituted by Ala"

XX  
PN WO9940198-A2.

PD 12-AUG-1999.

PF 04-FEB-1999; 99WO-EP00748.

PR 06-FEB-1998; 98EP-0200365.

PR 04-FEB-1998; 98EP-0200323.

XX  
PA (COLL/) COLLEN D J.

PA (LEUV-) LEUVEN RES & DEV VZW.

XX Collen DJ;  
XX WPI; 1999-508504/42.  
XX Staphylokinase derivatives with reduced immunogenicity, used for,  
PT e.g. treatment of arterial thrombosis  
XX  
PS Claim 7; Page -; 101pp; English.  
XX The present sequence is a specifically claimed Staphylokinase SakSTAR  
CC variant. This variant has one aminoacid that has been substituted  
CC by another aminoacid that reduces the reactivity with monoclonal  
CC antibodies and absorption of SakSTAR-specific antibodies from plasma  
CC of patients treated with staphylokinase. The derivatives can also be  
CC substituted with cysteine modified with PEG to maintain the specific  
CC activity and significantly reduce the plasma clearance. They have  
CC altered immunogenicity without markedly reducing the specific activity.  
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic  
CC potency in human plasma. The new staphylokinase derivatives are used  
CC for treatment of arterial thrombosis, especially myocardial infarction.  
CC The compositions can be used in human or veterinary practice.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the Staphylococcus aureus wild type staphylokinase sequence  
CC given in figure 1.  
XX  
SQ Sequence 136 AA;  
  
Query Match 98.9%; Score 707; DB 20; Length 136;  
Best Local Similarity 99.3%; Pred. No. 1.4e-69;  
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 SSSFDKGGKKGGDDASYFEPTGPYLMVNTGVDSKGNELLSPHYVEFPKPGTTLTKEKI 60  
Db 1 sssfdkgykkgddasyfepctgpylmvntgvdskgnellspkyvefpikpgtlttkeki 60  
  
QY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTYVDKNNKKKEETKSFPITEKGFVVPDLSEHI 120  
Db 61 eyvewaldataykefrvvelapsakievtlydknkkkeetksfptekgfvpdlsehi 120  
  
QY 121 KNPGFNLTITKVIEKK 136  
Db 121 knpgfnltitkviekk 136  
  
RESULT 32  
AAW44689  
ID AAW44689 standard; protein; 136 AA.  
XX  
AC AAW44689;  
XX  
DT 11-MAY-1998 (first entry)  
XX  
DE Staphylokinase mutant (D41N), useful as a thrombolytic agent.  
XX  
KW Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.  
XX  
OS Staphylococcus aureus.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 14 /note- "Wild-type Asp has been replaced by Asn"  
FT  
PN JP10028587-A.  
XX  
PD 03-FEB-1998.  
XX  
PF 19-JUL-1996; 96JP-0208991.  
XX  
PR 19-JUL-1996; 96JP-0208991.  
XX

PA (HONS ) YAKULT HONSHA KK.  
XX  
XX WPI; 1998-162525/15.  
XX  
XX Staphylokinase mutants - useful as thrombolytic agents  
PT  
PS Claim 1; Pages 5-6; 9pp; Japanese.  
XX  
XX The present sequence represents a specifically claimed staphylokinase  
CC (SAK) mutant (D41N). The invention relates to seven staphylokinase  
CC mutants which are useful as thrombolytic agents. The mutants are: (2) a  
CC SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136  
CC amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th  
CC Glu in SAK is replaced by Gln; (4) a SAK mutant (K113A) in which the 86th  
CC Lys in SAK is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th  
CC Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th  
CC Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th  
CC Asp in SAK is replaced by Glu, and (8) a SAK mutant (K136A) in which the  
CC 109th Lys in SAK is replaced by Ala.  
XX  
SQ Sequence 136 AA;  
  
Query Match 98.7%; Score 706; DB 19; Length 136;  
Best Local Similarity 98.5%; Pred. No. 1.8e-69;  
Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 SSSFDKGGKKGGDDASYFEPTGPYLMVNTGVDSKGNELLSPHYVEFPKPGTTLTKEKI 60  
Db 1 sssfdkgykkgddasyfepctgpylmvntgvdskgnellspkyvefpikpgtlttkeki 60  
  
QY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTYVDKNNKKKEETKSFPITEKGFVVPDLSEHI 120  
Db 61 eyvewaldataykefrvvelapsakievtlydknkkkeetksfptekgfvpdlsehi 120  
  
QY 121 KNPGFNLTITKVIEKK 136  
Db 121 knpgfnltitkviekk 136  
  
RESULT 33  
AAW44692  
ID AAW44692 standard; protein; 136 AA.  
XX  
AC AAW44692;  
XX  
DT 11-MAY-1998 (first entry)  
XX  
DE Staphylokinase mutant (K38S), useful as a thrombolytic agent.  
XX  
KW Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.  
XX  
OS Staphylococcus aureus.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 11 /note- "Wild-type Lys has been replaced by Ser"  
FT  
PN JP10028587-A.  
XX  
PD 03-FEB-1998.  
XX  
PF 19-JUL-1996; 96JP-0108991.  
XX  
PR 19-JUL-1996; 96JP-0208991.  
XX  
PA (HONS ) YAKULT HONSHA KK.  
XX  
XX WPI; 1998-162525/15.  
XX  
PT Staphylokinase mutants - useful as thrombolytic agents  
XX

PS Claim 1; Page 7; 9pp; Japanese.

XX  
CC The present sequence represents a specifically claimed staphylokinase  
CC (SAK) mutant (K38S). The invention relates to seven staphylokinase  
CC mutants which are useful as thrombolytic agents. The mutants are: (2) a  
CC SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136  
CC amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th  
CC Glu in SAK is replaced by Gln; (4) a SAK mutant (K113A) in which the 86th  
CC Lys in SAK is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th  
CC Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th  
CC Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th  
CC Asp in SAK is replaced by Glu, and (8) a SAK mutant (K136A) in which the  
CC 109th Lys in SAK is replaced by Ala.

XX  
SQ Sequence 136 AA;

Query Match 98.7%; Score 706; DB 19; Length 136;  
Best Local Similarity 98.5%; Pred. No. 1.8e-69;  
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SSSFDKGRYKGGDDASYFEPTGPYLMVNTGVDSKGNELSPHYVEFPPIKPGTTLTKEKI 60  
Db 1 sssfdkgyksgddasyfepptgpylmvntgvdskgnellsphyvefpikpgtlttkeki 60

OY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
Db 61 eyvewaldataykefrveldp sakievtydydknkkkeetksfpitekgfvvpdlsehi 120

OY 121 KNPGENLITKVIEKK 136  
Db 121 knpgfnlittkviekk 136

RESULT 34  
AAV15028  
ID AAV15028 standard; Protein; 136 AA.

XX  
AC AAV15028;

XX  
DT 03-NOV-1999 (first entry)

XX  
DE Staphylokinase variant SakSTAR (K130T, K135R).

XX  
KW Staphylokinase; immunogenic; specificity; SakSTAR (K130T, K135R);  
KW variant; derivative; anti-thrombotic; fibrinolytic; cardiant; veterinary;  
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis; PEG;  
KW monoclonal antibody; cysteine; polyethylene glycol; plasma clearance.

XX  
OS Staphylococcus aureus.

XX  
OS Synthetic.

FH Key Location/Qualifiers  
FT Misc-difference 130 /note= "Wild type Lys is substituted by Thr"  
FT Misc-difference 135 /note= "Wild type Lys is substituted by Arg"

XX  
PN WO9940198-A2.

XX  
PD 12-AUG-1999.

XX  
PF 04-FEB-1999; 99WO-EP00748.

XX  
PR 06-FEB-1998; 98EP-0200365.  
PR 04-FEB-1998; 98EP-0200323.

XX  
PA (COLL/) COLLEN D J.  
PA (LEUV-) LEUVEN RES & DEV VZW.

XX  
PI Collen DJ;

XX  
DR WPI; 1999-508504/42.

XX  
PT Staphylokinase derivatives with reduced immunogenicity, used for,  
PT e.g. treatment of arterial thrombosis

XX  
PS Claim 7; Page -; 101pp; English.

XX  
CC The present sequence is a specifically claimed staphylokinase SakSTAR  
CC variant. This variant has more than one aminoacid that has been  
CC substituted by another aminoacid that reduces the reactivity with  
CC monoclonal antibodies and absorption of SakSTAR-specific antibodies from  
CC plasma of patients treated with staphylokinase. The derivatives can also  
CC be substituted with cysteine modified with PEG to maintain the specific  
CC activity and significantly reduce the plasma clearance. They have  
CC altered immunogenicity without markedly reducing the specific activity.  
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic  
CC potency in human plasma. The new staphylokinase derivatives are used for  
CC treatment of arterial thrombosis, especially myocardial infarction. The  
CC compositions can be used in human or veterinary practice.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the Staphylococcus aureus wild type staphylokinase sequence  
CC given in figure 1.

XX  
SQ Sequence 136 AA;

Query Match 98.7%; Score 706; DB 20; Length 136;  
Best Local Similarity 98.5%; Pred. No. 1.8e-69;  
Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSSFDKGRYKGGDDASYFEPTGPYLMVNTGVDSKGNELSPHYVEFPPIKPGTTLTKEKI 60  
Db 1 sssfdkgyksgddasyfepptgpylmvntgvdskgnellsphyvefpikpgtlttkeki 60

OY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
Db 61 eyvewaldataykefrveldp sakievtydydknkkkeetksfpitekgfvvpdlsehi 120

OY 121 KNPGENLITKVIEKK 136  
Db 121 knpgfnlittkviekk 136

RESULT 35  
AAV15014  
ID AAV15014 standard; Protein; 136 AA.

XX  
AC AAV15014;

XX  
DT 03-NOV-1999 (first entry)

XX  
DE Staphylokinase variant SakSTAR (Y92A).

XX  
KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;  
KW SakSTAR (Y92A); anti-thrombotic; fibrinolytic; cardiant; veterinary;  
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;  
KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.

XX  
OS Staphylococcus aureus.

XX  
OS Synthetic.

FH Key Location/Qualifiers  
FT Misc-difference 92 /note= "Wild type Tyr is substituted by Ala"

XX  
PN WO9940198-A2.

XX  
PD 12-AUG-1999.

XX  
PF 04-FEB-1999; 99WO-EP00748.

XX  
PR 06-FEB-1998; 98EP-0200365.  
PR 04-FEB-1998; 98EP-0200323.

PA	(COLL/)	COLLEN D J.	
PA	(LEUV-)	LEUVEN RES & DEV VZW.	
XX			
PI		Collen DJ;	
XX			
DR	WPI;	1999-508504/42.	
XX			
PT	Staphylokinase derivatives with reduced immunogenicity, used for,		
PT	e.g. treatment of arterial thrombosis		
XX			
PS	Claim 7; Page -; 101pp; English.		
XX			
CC	The present sequence is a specifically claimed Staphylokinase SakSTAR		
CC	variant. This variant has one aminoacid that has been substituted		
CC	by another aminoacid that reduces the reactivity with monoclonal		
CC	antibodies and absorption of SakSTAR-specific antibodies from plasma		
CC	of patients treated with staphylokinase. The derivatives can also be		
CC	substituted with cysteine modified with PEG to maintain the specific		
CC	activity and significantly reduce the plasma clearance. They have		
CC	altered immunogenicity without markedly reducing the specific activity.		
CC	This sequence has anti-thrombotic, cardiant activity and a fibrinolytic		
CC	potency in human plasma. The new staphylokinase derivatives are used		
CC	for treatment of arterial thrombosis, especially myocardial infarction.		
CC	The compositions can be used in human or veterinary practice.		
CC	Note: The present sequence is not shown in the specification, but is		
CC	derived from the Staphylococcus aureus wild type staphylokinase sequence		
CC	given in figure 1.		
XX			
SQ	Sequence	136 AA;	
	Query Match	98.7%;	Score 706; DB 20; Length 136;
	Best Local Similarity	99.3%;	Pred. No. 1.8e-69;
	Matches 135; Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
OY	1	SSSEFDKGGKKKDDASYEEPTGPYLMVNVTVGDSKGNELSPHYVEFPPIKPGTTLTKKEKI	60
DB	1	sssfdkgkykkddasyfepcpgylmrvntvgvdsksgnelssphyvefpikpgtlttkel	60
OY	61	EYVWEALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSPITEKGFVVPDLSEHI	120
DB	61	eyvwealdataykefrvveldpksakievtyadknkkkeetksfptekgfvvpdlseh	120
OY	121	KNPGFNLTITKVIEKK	136
DB	121	knpgfnlittkvlekk	136
	RESULT	36	
	AAB01297		
ID	AAB01297	standard; protein; 136 AA.	
XX			
XX	AAB01297;		
AC			
DT	25-SEP-2000	(first entry)	
XX			
DE	Wild type staphylokinase from Staphylococcus aureus.		
XX			
KW	Immunogenicity; immunogen; T cell epitope; T-lymphocyte; drug;		
KW	vaccine; carrier; diagnosis; therapy; complement; C3; C5; CVF;		
KW	plasmnogen; staphylokinase; fibrin; blood clot; thrombolysis;		
KW	plasmn; myocardial infarction.		
XX			
OS	Staphylococcus aureus.		
XX			
PN	WO200034317-A2.		
XX			
PD	15-JUN-2000.		
XX			
PF	08-DEC-1999;	99WO-GB04119.	
XX			
PR	08-DEC-1998;	98GB-0026925.	
PR	02-FEB-1999;	99GB-0002139.	

XX	(BIOV-) BIOVATION LTD.
PA	Carr FJ, Adair FS, Hamilton AA, Carter G;
XX	WPI; 2000-423372/36.
XX	
DR	
XX	
PT	Rendering a protein non-immunogenic or less immunogenic useful in
PT	medicine and in diagnostics involves determining the amino acid
PT	sequence of the protein, identifying and modifying potential epitopes
PS	Example 3; Fig 7; 42pp; English.
XX	
CC	Proteins or their fragments can be rendered non-immunogenic or less
CC	immunogenic by identifying one or more potential T cell epitopes and
CC	modifying the sequence to eliminate at least one of these T cell
CC	epitopes to reduce the immunogenicity of the protein when exposed to
CC	the immune system of another organism. The method can be used for
CC	producing a protein with an enzymatic activity which has a beneficial
CC	therapeutic effect, a protein used to convert inactive drugs to its
CC	active form within a living organism, a vaccine, a protein used as a
CC	carrier of other molecule or a protein which binds to other molecules
CC	within or introduced within the living organism in order to alter the
CC	bio distribution of other molecules, such that the protein does not
CC	stimulate immune response in the living organism. The less immunogenic
CC	protein is useful in medicine, diagnosis and in manufacture of a
CC	therapeutic or a diagnostic agent. Staphylokinase is produced by
CC	Staphylococcus aureus. The protein holds considerable promise for
CC	thrombolytic therapy. Clinical use in humans has been limited due to
CC	the native protein being immunogenic. Availability of a
CC	non-immunogenic staphylokinase would have considerable importance as
CC	a potential agent for thrombolytic therapy. This is the sequence
CC	of the wild type staphylokinase. The altered staphylokinase sequence
CC	is given in AAB01298. See GENESeq records AAB01289-B01302.
XX	
SQ	Sequence 136 AA;
Query Match	98.7%; Score 706; DB 21; Length 136;
Best Local Similarity	99.3%; Pred. No. 1.8e-69;
Matches 135; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
OY	1 SSSFDKGGKKKGDDASYFEPTGPLYMNVNTGVDSKGNELISPHYVEFPKPGTTLTKREKI 60 
Db	1 sssfdkgkykkgddasyfepctpylmvntgydskgnellsphavefpikpgtlttkel 60 
OY	61 EYYEWALDATAYKEFRVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFVPDLSEHI 120 
Db	61 eyyewaldataykefrvelgpsakievtydydnknkkkeetksfpitekfvpdlsehl 120 
OY	121 KNPGFNLTKVIEKK 136 
Db	121 knpgfnlitzkviekk 136 
RESULT 37	
AAMW44691	
ID	AAMW44691 standard; protein; 136 AA.
AC	AAMW44691;
XX	
XX	11-MAR-1998 (first entry)
DT	
XX	
DE	Staphylokinase mutant (K113A), useful as a thrombolytic agent.
XX	
KW	Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.
XX	
OS	Staphylococcus aureus.
OS	Synthetic.
XX	
FH	Key Location/Qualifiers
FT	Modified-site 86
FT	/note= "wild-type lys has been replaced by Ala"



```
XX PN JP10028587-A.
XX PD 03-FEB-1998.
XX PE 19-JUL-1996; 96JP-0108991.
XX PR 19-JUL-1996; 96JP-0208991.
XX PA (HONS ) YAKULT HONSHA KK.
XX DR WPI; 1998-162525/15.
XX PT Staphylokinase mutants - useful as thrombolytic agents
XX PS Claim 1; Pages 6-7; 9pp; Japanese.
XX SS The present sequence represents a specifically claimed staphylokinase
CC (SAK) mutant (K113A). The invention relates to seven staphylokinase
CC mutants which are useful as thrombolytic agents. The mutants are: (2) a
CC SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136
CC amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th
CC Glu in SAK is replaced by Gln; (4) a SAK mutant (K113A) in which the 86th
CC Lys in SAK is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th
CC Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th
CC Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th
CC Asp in SAK is replaced by Glu, and (8) a SAK mutant (K136A) in which the
CC 109th Lys in SAK is replaced by Ala.
XX SQ Sequence 136 AA;

Query Match 98.6%; Score 705; DB 19; Length 136;
Best Local Similarity 98.5%; Pred. No. 2.4e-69;
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSSFDKGGKKGGDASYPEPTGPLYMNVNTGVDKSGNELSPHYVEFPKPGTTLTKRKI 60
   |||||||
Db 1 sssfdkgkykkgdasyfepptgplymwnvtgvdgkgnellsphvvefpikpgtltlkeki 60

QY 61 EYVWEALDAPYKKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120
   |||||||
Db 61 eyvewaldataykefrvvelpsaaievtlydknkkkeetksfpitekgfvvpdlsehi 120

QY 121 KNPGFNLTIKVIEKK 136
   |||||||
Db 121 knpgfnliltkvielek 136

RESULT 38
AAW44695
ID AAW44695 standard; protein; 136 AA.
XX AC AAW44695;
XX DT 11-MAY-1998 (first entry)
XX DE Staphylokinase mutant (K136A), useful as a thrombolytic agent.
XX KW Staphylokinase mutant; thrombolytic agent; Staphylococcus aureus.
XX OS Staphylococcus aureus.
XX OS Synthetic.
XX FH Key
FT Modified-site 109 Location/Qualifiers
FT /note= "Wild-type Lys has been replaced by Ala"
XX PN JP10028587-A.
XX PD 03-FEB-1998.
XX PF 19-JUL-1996; 96JP-0108991.
```

```
XX PR 19-JUL-1996; 96JP-0208991.
XX PA (HONS ) YAKULT HONSHA KK.
XX DR WPI; 1998-162525/15.
XX PT Staphylokinase mutants - useful as thrombolytic agents
XX PS Claim 1; Page 8; 9pp; Japanese.
XX SS The present sequence represents a specifically claimed staphylokinase
CC (SAK) mutant (K136A). The invention relates to seven staphylokinase
CC mutants which are useful as thrombolytic agents. The mutants are: (2) a
CC SAK mutant (D41N) in which the 14th Asp in the SAK which comprises 136
CC amino acids is replaced by Asn; (3) a SAK mutant (E85Q) in which the 58th
CC Glu in SAK is replaced by Gln; (4) a SAK mutant (K113A) in which the 86th
CC Lys in SAK is replaced by Ala; (5) a SAK mutant (K38S) in which the 11th
CC Lys in SAK is replaced by Ser; (6) a SAK mutant (E73Q) in which the 46th
CC Glu in SAK is replaced by Gln; (7) a SAK mutant (D96E) in which the 69th
CC Asp in SAK is replaced by Glu, and (8) a SAK mutant (K136A) in which the
CC 109th Lys in SAK is replaced by Ala.
XX SQ Sequence 136 AA;

Query Match 98.6%; Score 705; DB 19; Length 136;
Best Local Similarity 98.5%; Pred. No. 2.4e-69;
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSSFDKGGKKGGDASYPEPTGPLYMNVNTGVDKSGNELSPHYVEFPKPGTTLTKRKI 60
   |||||||
Db 1 sssfdkgkykkgdasyfepptgplymwnvtgvdgkgnellsphvvefpikpgtltlkeki 60

QY 61 EYVWEALDAPYKKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120
   |||||||
Db 61 eyvewaldataykefrvvelpsakievtlydknkkkeetksfpiteagfvvpdlsehi 120

QY 121 KNPGFNLTIKVIEKK 136
   |||||||
Db 121 knpgfnliltkvielek 136

RESULT 39
AAV15041
ID AAV15041 standard; Protein; 136 AA.
XX AC AAV15041;
XX DT 03-NOV-1999 (first entry)
XX DE Staphylokinase variant SakSTAR (K74Q, K130E, K135R).
XX KW Staphylokinase; immunogenic; specificity; SakSTAR (K74Q, K130E, K135R);
XX KW variant; derivative; anti-thrombotic; fibrinolytic; cardiant; veterinary;
XX KW Staphylococcus aureus; myocardial infarction; arterial thrombosis; PEG;
XX KW monoclonal antibody; cysteine; polyethylene glycol; plasma clearance.
XX OS Staphylococcus aureus.
XX OS Synthetic.
XX FH Key
FT Misc-difference 74 Location/Qualifiers
FT /note= "Wild type Lys is substituted by Gln"
FT Misc-difference 130 /note= "Wild type Lys is substituted by Glu"
FT Misc-difference 135 /note= "Wild type Lys is substituted by Arg"
XX PN WO9940198-A2.
XX PD 12-AUG-1999.
```

PF 04-FEB-1999; 99WO-EP00748.  
XX  
PR 06-FEB-1998; 98EP-0200365.  
PR 04-FEB-1998; 98EP-0200323.  
XX  
PA (COLL/) COLLEN D J.  
PA (LEUV-) LEUVEN RES & DEV VZW.  
XX  
XX Collen DJ;  
PI  
XX  
DR WPI; 1999-508504/42.  
XX  
PT Staphylokinase derivatives with reduced immunogenicity, used for,  
PT e.g. treatment of arterial thrombosis  
PS  
XX Claim 7; Page -; 101pp; English.  
XX  
CC The present sequence is a specifically claimed Staphylokinase SakSTAR  
CC variant. This variant has more than one aminoacid that has been  
CC substituted by another aminoacid that reduces the reactivity with  
CC monoclonal antibodies and absorption of SakSTAR-specific antibodies from  
CC plasma of patients treated with staphylokinase. The derivatives can also  
CC be substituted with cysteine modified with PEG to maintain the specific  
CC activity and significantly reduce the plasma clearance. They have  
CC altered immunogenicity without markedly reducing the specific activity.  
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic  
CC potency in human plasma. The new staphylokinase derivatives are used for  
CC treatment of arterial thrombosis, especially myocardial infarction. The  
CC compositions can be used in human or veterinary practice.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the Staphylococcus aureus wild type staphylokinase sequence  
CC given in figure 1.  
XX  
SQ Sequence 136 AA;  
  
Query Match 98.5%; Score 704; DB 20; Length 136;  
Best Local Similarity 97.8%; Pred. No. 3e-69;  
Matches 133; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SSSEFDKGYKKGGDDASYFEPTGYLMVNVGTGVDKGNELLSPHYVEFPPIKPGTTLTKEKI 60  
Db 1 sssfdkgykkggdaasyfep tgy lmvnvtgvdskgnellsp hyvefpikpgt tltkeki 60  
  
QY 61 EYYEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
Db 61 eyyewaldata ykefrvvel dpsakievt yydknkkkeetkspf itekgf vvpdlsehi 120  
  
QY 121 KNPGFNLTQVIEKK 136  
Db 121 knpgfnltqv iekk 136  
  
RESULT 40  
ID AAY15020 standard; Protein; 136 AA.  
XX  
AC AAY15020;  
XX  
DT 03-NOV-1999 (first entry)  
XX  
DE Staphylokinase variant SakSTAR (G36R, K74R).  
XX  
KW Staphylokinase; variant; immunogenic; specificity; derivative; cysteine;  
KW SakSTAR (G36R, K74R); anti-thrombotic; fibrinolytic; cardiant;  
KW Staphylococcus aureus; myocardial infarction; arterial thrombosis;  
KW monoclonal antibody; polyethylene glycol; PEG; plasma clearance.  
XX  
OS Staphylococcus aureus.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 36

FT /note- "wild type Gly is substituted by Arg"  
FT Misc-difference 74  
FT /note- "wild type Lys is substituted by Arg"  
XX  
XX WO9940198-A2.  
PN  
XX  
PD 12-AUG-1999.  
XX  
XX  
PE 04-FEB-1999; 99WO-EP00748.  
XX  
XX  
PR 06-FEB-1998; 98EP-0200365.  
PR 04-FEB-1998; 98EP-0200323.  
XX  
XX  
PA (COLL/) COLLEN D J.  
PA (LEUV-) LEUVEN RES & DEV VZW.  
XX  
XX Collen DJ;  
PI  
XX  
DR WPI; 1999-508504/42.  
XX  
XX  
PT Staphylokinase derivatives with reduced immunogenicity, used for,  
PT e.g. treatment of arterial thrombosis  
PS  
XX Claim 7; Page -; 101pp; English.  
XX  
CC The present sequence is a specifically claimed Staphylokinase SakSTAR  
CC variant. This variant has more than one aminoacid that has been  
CC substituted by another aminoacid that reduces the reactivity with  
CC monoclonal antibodies and absorption of SakSTAR-specific antibodies from  
CC plasma of patients treated with staphylokinase. The derivatives can also  
CC be substituted with cysteine modified with PEG to maintain the specific  
CC activity and significantly reduce the plasma clearance. They have  
CC altered immunogenicity without markedly reducing the specific activity.  
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic  
CC potency in human plasma. The new staphylokinase derivatives are used for  
CC treatment of arterial thrombosis, especially myocardial infarction. The  
CC compositions can be used in human or veterinary practice.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the Staphylococcus aureus wild type staphylokinase sequence  
CC given in figure 1.  
XX  
SQ Sequence 136 AA;  
  
Query Match 98.5%; Score 704; DB 20; Length 136;  
Best Local Similarity 98.5%; Pred. No. 3e-69;  
Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 SSSEFDKGYKKGGDDASYFEPTGYLMVNVGTGVDKGNELLSPHYVEFPPIKPGTTLTKEKI 60  
Db 1 sssfdkgykkggdaasyfep tgy lmvnvtgvdskgnellsp hyvefpikpgt tltkeki 60  
  
QY 61 EYYEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
Db 61 eyyewaldata ykefrvvel dpsakievt yydknkkkeetkspf itekgf vvpdlsehi 120  
  
QY 121 KNPGFNLTQVIEKK 136  
Db 121 knpgfnltqv iekk 136  
  
RESULT 41  
ID AAY15012 standard; Protein; 136 AA.  
XX  
AC AAY15012;  
XX  
DT 03-NOV-1999 (first entry)  
XX  
DE Staphylokinase variant SakSTAR (D82A, S84A).  
XX  
KW Staphylokinase; variant; immunogenic; specificity; SakSTAR (D82A, S84A);  
KW derivative; anti-thrombotic; fibrinolytic; cardiant; veterinary;  
XX

```

KW Staphylococcus aureus; myocardial infarction; arterial thrombosis; PEG;
KM monoclonal antibody; cysteine; polyethylene glycol; plasma clearance.
XX
OS Staphylococcus aureus.
XX Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 82 /note= "Wild type Asp is substituted by Ala"
FT Misc-difference 84 /note= "Wild type Ser is substituted by Ala"
FT
XX
PN W09940198-A2.
XX
PD 12-ADG-1999.
XX
PF 04-FEB-1999; 99WO-EP00748.
XX
PR 06-FEB-1998; 98EP-0200365.
PR 04-FEB-1998; 98EP-0200323.
XX
PA (COLL/) COLLEN D J.
PA (LEUV-) LEUVEN RES & DEV VZW.
XX
PI Collen DJ;
XX
DR WPI; 1999-508504/42.
XX
XX
XX Staphylokinase derivatives with reduced immunogenicity, used for,
PT e.g. treatment of arterial thrombosis
XX
XX Claim 7; Page -: 101pp; English.
XX
XX The present sequence is a specifically claimed Staphylokinase SakSTAR
CC variant. This variant has more than one aminoacid that has been
CC substituted by another aminoacid that reduces the reactivity with
CC monoclonal antibodies and absorption of SakSTAR-specific antibodies from
CC plasma of patients treated with staphylokinase. The derivatives can also
CC be substituted with cysteine modified with PEG to maintain the specific
CC activity and significantly reduce the plasma clearance. They have
CC altered immunogenicity without markedly reducing the specific activity.
CC This sequence has anti-thrombotic, cardiant activity and a fibrinolytic
CC potency in human plasma. The new staphylokinase derivatives are used for
CC treatment of arterial thrombosis, especially myocardial infarction. The
CC compositions can be used in human or veterinary practice.
CC Note: The present sequence is not shown in the specification, but is
CC derived from the staphylococcus aureus wild type staphylokinase sequence
CC given in figure 1.
CC
XX
SQ Sequence 136 AA;
XX
XX
Query Match 98.5%; Score 704; DB 20; Length 136;
Best Local Similarity 98.5%; Pred. No. 3e-69;
Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 SSSFDKGYKKGGDDASYFEPTGPLYMWNVTGVDKSGNELSPHYVEFPPIKGTTLTKEKI 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1 sssfdkgykkgddasyfeptgplymwnvtgvdskgnellsphyvefpikpgttltkeki 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 61 EYVVEALDATAYKKEFRVVELDPSAKIEVTYYDDKNKKKEETKSPITEKGFWVPDLSEHI 120
Db ||||||||||||||||||| : ||||||||||||||||||||||||||||||||
QY 61 eyvvealdataykefrvelapaaaklevtydknkkkeetksfptekgfvvpdlsehi 120
Db ||||||||||||||||||| : ||||||||||||||||||||||||||||||||
QY 121 KNPGFNLTQVIEKK 136
Db |||||||||||||||||||
QY 121 knpgfnltkvviekk 136
Db |||||||||||||||||||
XX
RESULT 42
ID AAW03079 standard; protein; 136 AA.
XX

```

AC	AAW03079;	
XX	19-FEB-1997	(first entry)
XX	Staphylokinase derivative M10 with altered charge cluster 10.	
DE	Staphylokinase; mutant; mutein; variant; immunogenicity; decrease;	
XX	derivative; SakSTAR; arterial thrombosis; thrombolytic agent.	
KW	Staphylococcus aureus.	
XX	Synthetic.	
OS		
XX		
FH	Key	Location/Qualifiers
FT	Region	86..88
FT		/label= altered_M10_cluster
FT		/note= "wild-type cluster 10 sequence Lys-Ile-Glu
FT		has been replaced by Ala-Ile-Ala to reduce
FT		immunogenicity"
XX		
PN	EP721982-A1.	
PD	17-JUL-1996.	
XX		
PF	06-JAN-1995;	95EP-0200023.
XX		
PR	06-JAN-1995;	95EP-0200023.
XX		
PA	(COLL/) COLLEN D J.	
PA	(LEUV-) LEUVEN RES & DEV VZW.	
XX		
PI	Collen D;	
DR	WPI; 1996-322832/33.	
XX		
PT	New staphylokinase mutants with reduced immunogenicity - useful for	
PT	treating arterial thrombosis	
XX		
PS	Example 2; Page -; 21pp; English.	
XX		
CC	Staphylokinase derivatives showing a reduced immunogenicity as	
CC	compared to wild-type staphylokinase are claimed. The derivatives	
CC	are useful as thrombolytic agents to treat arterial thrombosis and	
CC	are pref. produced by eliminating at least one of the epitopes	
CC	indicated in the features table. The epitopes are destroyed by	
CC	replacing one or more amino acid residues in a charge cluster by an	
CC	Ala residue. Mutations are introduced using site-directed	
CC	mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus	
CC	lysogenic strain SakSTAR. The present sequence is a specific	
CC	example of a staphylokinase derivative according to the invention;	
CC	the sequence does not appear in the specification but can be generated	
CC	from the wild-type staphylokinase sequence using the description given	
XX	in Table 1 of the disclosure.	
SQ	Sequence 136 AA;	

```

Query Match          98.3%; Score 703; DB 17; Length 136;
Best Local Similarity 98.5%; Pred. No. 3.9e-69;
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 SSSFDKGYKKKGDDASYFEPTGPYLMVNVTVGVDSDSGNELSPHYVEPIKPGTTLTKEKI 60
        |||
Db       1 sssfdkgykkkgddasyfeptgpylmvntvgvdskgnellsphyvefpikpgttltkeki 60

QY      61 EYYVEWALDATAYPEFRVELDPASAKIEVTYYDDKNKKKEETKSFPITEKGFVPDLSEHI 120
        |||
Db       61 eyyvewaldataypefrveldpasaiaavtydydknkkkeetksfptekgfvvpdlsehi 120

QY      121 KNPGEFLTTRKVIEKK 136
        |||
Db       121 knpgfniltkvielekk 136

```

RESULT	43
ID	AAW03078 standard; protein; 136 AA.
AC	AAW03078;
DT	19-FEB-1997 (first entry)
XX	
DE	Staphylokinase derivative M6 with altered charge cluster 6.
XX	
KW	Staphylokinase; mutant; mutein; variant; immunogenicity; decrease; derivative; SakSTAR; arterial thrombosis; thrombolytic agent.
XX	
OS	Staphylococcus aureus. Synthetic.
FH	
FT	Key Location/Qualifiers Region 61..65 /label= altered_M6_cluster /note= "wild-type cluster 6 sequence Glu-Tyr-Tyr-Val-Glu has been replaced by Ala-Tyr-Tyr-Val-Ala to reduce immunogenicity"
XX	
PN	EP721982-A1.
XX	
PD	17-JUL-1996.
XX	
Pf	06-JAN-1995; 95EP-0200023.
XX	
PR	06-JAN-1995; 95EP-0200023.
PA	(COLL/) COLLEN D J. (LEUV-) LEUVEN RES & DEV VZR.
PI	Collen D;
DR	WPI; 1996-322832/33.
PT	New staphylokinase mutants with reduced immunogenicity - useful for treating arterial thrombosis
PS	Example 2; Page -: 21pp; English.
CC	
CC	Staphylokinase derivatives showing a reduced immunogenicity as compared to wild-type staphylokinase are claimed. The derivatives are useful as thrombolytic agents to treat arterial thrombosis and are pref. produced by eliminating at least one of the epitopes indicated in the features table. The epitopes are destroyed by replacing one or more amino acid residues in a charge cluster by an Ala residue. Mutations are introduced using site-directed mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus lysogenic strain SakSTAR. The present sequence is a specific example of a staphylokinase derivative according to the invention; the sequence does not appear in the specification but can be generated from the wild-type staphylokinase sequence using the description given in Table 1 of the disclosure.
SO	Sequence 136 AA;
QY	Query Match 98.3%; Score 703; DB 17; Length 136; Best Local Similarity 98.5%; Pred. No. 3.9e-69; Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
DB	1 SSFDPKGGKKGGDDASYFEPTGPILMVNTGVDSKGNELSPHYVEFPRIKPGTTLTKKEI 60       1 sssfdkgkykgyddasyfeptgpylmvntgvdsksgnellsphyvefpikpgtlttkeki 60
QY	61 EYYVEMALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFFITEKGFPVPDLSEHI 120       61 ayyvawaldataykefrvveldpsakievtlydknkkkkeetksffitekgtfvvpdlsehi 120
OY	121 KNDGFNLITKVIEKK 136

[illegible]



|||||  
Db 61 eyyewaldataykefrvvelpsakievtydknkkkeetksfptaagfvvpdlsehl 120  
QY 121 KNPGENLITKVIIEKK 136  
|||||  
Db 121 knpgfnlltkvielkk 136

RESULT 45  
AAW03076  
ID AAW03076 standard; protein; 136 AA.  
AC AAW03076;  
XX  
DT 19-FEB-1997 (first entry)  
XX  
DE Staphylokinase derivative M21 with altered charge cluster 21.  
XX  
KW Staphylokinase; mutant; mutein; variant; immunogenicity; decrease;  
KW derivative; SakSTAR; arterial thrombosis; thrombolytic agent.  
XX  
OS Staphylococcus aureus.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Region 8..10  
FT /label= altered\_M21\_cluster  
FT /note= "wild-type cluster 21 sequence Lys-Tyr-Lys  
FT has been replaced by Ala-Tyr-Ala to  
FT reduce immunogenicity"  
XX  
PN EP721982-A1.  
XX  
PD 17-JUL-1996.  
XX  
PF 06-JAN-1995; 95EP-0200023.  
XX  
PR 06-JAN-1995; 95EP-0200023.  
XX  
PA (COLL/) COLLEN D J.  
PA (LEUV-) LEUVEN RES & DEV VZW.  
XX  
PI Collen D;  
XX  
DR WPI; 1996-322832/33.  
XX  
PT New staphylokinase mutants with reduced immunogenicity - useful for  
PT treating arterial thrombosis  
XX  
PS Example 2; Page -: 21pp; English.  
XX  
CC Staphylokinase derivatives showing a reduced immunogenicity as  
CC compared to wild-type staphylokinase are claimed. The derivatives  
CC are useful as thrombolytic agents to treat arterial thrombosis and  
CC are pref. produced by eliminating at least one of the epitopes  
CC indicated in the features table. The epitopes are destroyed by  
CC replacing one or more amino acid residues in a charge cluster by an  
CC Ala residue. Mutations are introduced using site-directed  
CC mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus  
CC lysogenic strain SakSTAR. The present sequence is a specific  
CC example of a staphylokinase derivative according to the invention;  
CC the sequence does not appear in the specification but can be generated  
CC from the wild-type staphylokinase sequence using the description given  
CC in Table 1 of the disclosure.  
XX  
SQ Sequence 136 AA;

Query Match 98.3%; Score 703; DB 17; Length 136;  
Best Local Similarity 98.5%; Pred. No. 3.9e-69;  
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 SSSFdKGKYGKGDASYFEPTGPLYMNVNTGVDSKGNELLSPHYVEFPPIKPGTTLTKKEKI 60

|||||  
Db 1 sssfkdgayakgddasfyepptgplymnnvtgvdskgnellspHYVEFPPIKPGTTLTKKEKI 60  
QY 61 EYVEMALDATAYKKEFRVVELDPSAKIEVITYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
|||||  
Db 61 eyyewaldataykefrvvelpsakievtydknkkkeetksfptekgfvvpdlsehl 120

QY 121 KNPGENLITKVIIEKK 136  
|||||  
Db 121 knpgfnlltkvielkk 136

RESULT 46  
AAW03095  
ID AAW03095 standard; protein; 136 AA.  
AC AAW03095;  
XX  
DT 19-FEB-1997 (first entry)  
XX  
DE Staphylokinase derivative M10 with altered charge cluster 10.  
XX  
KW Staphylokinase; mutant; mutein; variant; immunogenicity; decrease;  
KW derivative; SakSTAR; arterial thrombosis; thrombolytic agent.  
XX  
OS Staphylococcus aureus.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Region 86..88  
FT /label= altered\_M10\_cluster  
FT /note= "wild-type cluster 10 sequence Lys-Ile-Glu  
FT has been replaced by Ala-Ile-Ala to reduce  
FT immunogenicity"  
XX  
PN W09621016-A2.  
XX  
PD 11-JUL-1996.  
XX  
PF 03-JAN-1996; 96WO-EP00081.  
XX  
PR 17-NOV-1995; 95JP-0299781.  
PR 06-JAN-1995; 95EP-0200023.  
PR 11-JAN-1995; 95US-0371505.  
PR 09-JUN-1995; 95EP-0201531.  
PR 06-JUL-1995; 95US-0499092.  
XX  
PA (COLL/) COLLEN D.  
PA (LEUV-) LEUVEN RES & DEV VZW.  
XX  
PI Collen D;  
XX  
DR WPI; 1996-333991/33.  
XX  
PT New staphylokinase derivs. having reduced immunogenicity - useful  
PT for treating arterial thrombosis  
XX  
PS Example 2; Page -: 58pp; English.  
XX  
CC Staphylokinase derivatives showing a reduced immunogenicity as  
CC compared to wild-type staphylokinase are claimed. The derivatives  
CC are useful as thrombolytic agents to treat arterial thrombosis and  
CC are pref. produced by eliminating at least one of the epitopes  
CC indicated in the features table. The epitopes are destroyed by  
CC replacing one or more amino acid residues in a charge cluster by an  
CC Ala residue. Mutations are introduced using site-directed  
CC mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus  
CC lysogenic strain SakSTAR. The present sequence is a specific  
CC example of a staphylokinase derivative according to the invention;  
CC the sequence does not appear in the specification but can be generated  
CC from the wild-type staphylokinase sequence using the description given  
CC in Table 1 of the disclosure.  
XX

SQ Sequence 136 AA;

Query Match 98.3%; Score 703; DB 17; Length 136;  
Best Local Similarity 98.5%; Pred. No. 3.9e-69;  
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSSFDKGGKYYKKGDDASYFEPTGPLYMNVNTGVDSKGNELSPHYVEFPPIKPGTTLTKEKI 60  
|||||  
Db 1 ssaefdkgkykkgddasyfeptgplymnnvtgvdskgnellspgyvefpikpgtlttkeki 60

QY 61 EYVEWALDATAKKEFRVVELDPSAKIEVTYYDKNKKKEETKSEPTTEKGFVVPDLSEHI 120  
|||||  
Db 61 eyvewaldataykefrvveldpsaaiavtydydknkkkeetksfptekgfvpdlsehi 120

QY 121 KNPGFNLITKRVIEKK 136  
|||||  
Db 121 knpgfnlittkviekk 136

RESULT 47  
AAW03098  
ID AAW03098 standard; protein; 136 AA.

XX AC AAW03098;  
XX DT 19-FEB-1997 (first entry)  
XX DE Staphylokinase derivative M15 with altered charge cluster 15.  
XX KW Staphylokinase; mutant; mutein; variant; immunogenicity; decrease;  
XX KW derivative; SakSTAR; arterial thrombosis; thrombolytic agent.  
XX OS Staphylococcus aureus.  
XX OS Synthetic.  
XX FH Key  
XX FT Region 108..109  
XX FT /label- altered\_M15-cluster  
XX FT /note- "wild-type cluster 15 sequence Glu-Lys  
XX FT has been replaced by Ala-Ala to reduce  
XX FT immunogenicity"  
XX PN W09621016-A2.  
XX PD 11-JUL-1996.  
XX PF 03-JAN-1996; 96WO-EP00081.  
XX PR 17-NOV-1995; 95JP-0299781.  
XX PR 06-JAN-1995; 95EP-0200023.  
XX PR 11-JAN-1995; 95US-0371505.  
XX PR 09-JUN-1995; 95EP-0201531.  
XX PR 06-JUL-1995; 95US-0499092.  
XX PA (COLL/) COLLEN D.  
XX PA (LEUV-) LEUVEN RES & DEV VZW.  
XX PI Colleen D;  
XX DR WPI; 1996-333991/33.  
XX PT New staphylokinase derivs. having reduced immunogenicity - useful  
XX PT for treating arterial thrombosis  
XX PS Example 2; Page -; 58pp; English.  
XX CC Staphylokinase derivatives showing a reduced immunogenicity as  
XX CC compared to wild-type staphylokinase are claimed. The derivatives  
XX CC are useful as thrombolytic agents to treat arterial thrombosis and  
XX CC are pref. produced by eliminating at least one of the epitopes  
XX CC indicated in the features table. The epitopes are destroyed by  
XX CC replacing one or more amino acid residues in a charge cluster by an

CC Ala residue. Mutations are introduced using site-directed  
CC mutagenesis on wild-type staphylokinase genomic DNA from the S. aureus  
CC lysogenic strain SakSTAR. The present sequence is a specific  
CC example of a staphylokinase derivative according to the invention;  
CC the sequence does not appear in the specification but can be generated  
CC from the wild-type staphylokinase sequence using the description given  
CC in Table 1 of the disclosure.

SQ Sequence 136 AA;

Query Match 98.3%; Score 703; DB 17; Length 136;  
Best Local Similarity 98.5%; Pred. No. 3.9e-69;  
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSSFDKGGKYYKKGDDASYFEPTGPLYMNVNTGVDSKGNELSPHYVEFPPIKPGTTLTKEKI 60  
|||||  
Db 1 ssaefdkgkykkgddasyfeptgplymnnvtgvdskgnellspgyvefpikpgtlttkeki 60

QY 61 EYVEWALDATAKKEFRVVELDPSAKIEVTYYDKNKKKEETKSEPTTEKGFVVPDLSEHI 120  
|||||  
Db 61 eyvewaldataykefrvveldpsakievtydydknkkkeetksfptaagfvvpdlsehi 120

QY 121 KNPGFNLITKRVIEKK 136  
|||||  
Db 121 knpgfnlittkviekk 136

RESULT 48  
AAW03073  
ID AAW03073 standard; protein; 136 AA.

XX AC AAW03073;  
XX DT 19-FEB-1997 (first entry)  
XX DE Staphylokinase derivative M3 with altered charge cluster 3.  
XX KW Staphylokinase; mutant; mutein; variant; immunogenicity; decrease;  
XX KW derivative; SakSTAR; arterial thrombosis; thrombolytic agent.  
XX OS Staphylococcus aureus.  
XX OS Synthetic.  
XX FH Key  
XX FT Region 35..38  
XX FT /label- altered\_M3-cluster  
XX FT /note- "wild-type cluster 3 sequence Lys-Gly-Asn-Glu  
XX FT has been replaced by Ala-Gly-Asn-Ala to  
XX FT reduce immunogenicity"  
XX PN EP721982-A1.  
XX PD 17-JUL-1996.  
XX PF 06-JAN-1995; 95EP-0200023.  
XX PR 06-JAN-1995; 95EP-0200023.  
XX PA (COLL/) COLLEN D J.  
XX PA (LEUV-) LEUVEN RES & DEV VZW.  
XX PI Colleen D;  
XX DR WPI; 1996-322832/33.  
XX PT New staphylokinase mutants with reduced immunogenicity - useful for  
XX PT treating arterial thrombosis  
XX PS Claim 9; Page -; 21pp; English.  
XX CC Staphylokinase derivatives showing a reduced immunogenicity as  
XX CC compared to wild-type staphylokinase are claimed. The derivatives

CC are useful as thrombolytic agents to treat arterial thrombosis and  
CC are pref. produced by eliminating at least one of the epitopes  
CC indicated in the features table. The epitopes are destroyed by  
CC replacing one or more amino acid residues in a charge cluster by an  
CC Ala residue. Mutations are introduced using site-directed  
CC mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus  
CC lysogenic strain SakSTAR. The present sequence is that of the  
CC specifically claimed staphylokinase derivative M3; the sequence  
CC does not appear in the specification but can be generated from the  
CC wild-type staphylokinase sequence using the description given in  
CC the claims.

XX Sequence 136 AA;

Query Match 98.3%; Score 703; DB 17; Length 136;  
Best Local Similarity 98.5%; Pred. No. 3.9e-69;  
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SSSFDKGYKKGGDDASYFEPTGPIYLMVNTGVDKGNELISPHYVEFPPIKPGTTLTKEKI 60  
|||||  
Db 1 sssfdkgykkggdaasyfptgpylmvntgvsagnalispkyvefpikpgtltlkeki 60

OY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTYVDKNNKKKEETKSPITEKGFVVPDLSEHI 120  
|||||  
Db 61 eyvewaldataykefrvvelpsakievtyvdnknkkeeetksfpitekgfvvpdlsehi 120

OY 121 KNPGENLITKRVIEKK 136  
|||||  
Db 121 knpgfnlitrkviekk 136

## RESULT 49

AAW03090 standard; protein; 136 AA.

XX AAW03090;

DT 19-FEB-1997 (first entry)

DE Staphylokinase derivative M3 with altered charge cluster 3.

KW Staphylokinase; mutant; mutein; variant; immunogenicity; decrease;  
derivative; SakSTAR; arterial thrombosis; thrombolytic agent.

OS Staphylococcus aureus.  
Synthetic.

XX Key Location/Qualifiers  
35.38

FT /label= altered\_M3\_cluster  
FT /note= "wild-type cluster 3 sequence Lys-Gly-Asn-Glu  
FT has been replaced by Ala-Gly-Asn-Ala to  
FT reduce immunogenicity"

PN WO9621016-A2.

PD 11-JUL-1996.

PF 03-JAN-1996; 96WO-EP00081.

PR 17-NOV-1995; 95JP-0299781.

PR 06-JAN-1995; 95EP-0200023.

PR 11-JAN-1995; 95US-0371505.

PR 09-JUN-1995; 95EP-0201531.

PA (COLL/) COLLEN D.  
PA (LEUV-) LEUVEN RES & DEV VZW.

PI Collen D;  
DR WPI; 1996-333991/33.

XX New staphylokinase derivs. having reduced immunogenicity - useful  
PT for treating arterial thrombosis

PS Claim 9; Page -: 58pp; English.

XX Staphylokinase derivatives showing a reduced immunogenicity as  
CC compared to wild-type staphylokinase are claimed. The derivatives  
CC are useful as thrombolytic agents to treat arterial thrombosis and  
CC are pref. produced by eliminating at least one of the epitopes  
CC indicated in the features table. The epitopes are destroyed by  
CC replacing one or more amino acid residues in a charge cluster by an  
CC Ala residue. Mutations are introduced using site-directed  
CC mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus  
CC lysogenic strain SakSTAR. The present sequence is that of the  
CC specifically claimed staphylokinase derivative M3; the sequence  
CC does not appear in the specification but can be generated from the  
CC wild-type staphylokinase sequence using the description given in  
CC the claims.

XX Sequence 136 AA;

Query Match 98.3%; Score 703; DB 17; Length 136;  
Best Local Similarity 98.5%; Pred. No. 3.9e-69;  
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SSSFDKGYKKGGDDASYFEPTGPIYLMVNTGVDKGNELISPHYVEFPPIKPGTTLTKEKI 60  
|||||  
Db 1 sssfdkgykkggdaasyfptgpylmvntgvsagnalispkyvefpikpgtltlkeki 60

OY 61 EYVWEALDATAYKEFRVVELDPSAKIEVTYVDKNNKKKEETKSPITEKGFVVPDLSEHI 120  
|||||  
Db 61 eyvewaldataykefrvvelpsakievtyvdnknkkeeetksfpitekgfvvpdlsehi 120

OY 121 KNPGENLITKRVIEKK 136  
|||||  
Db 121 knpgfnlitrkviekk 136

## RESULT 50

AAW03088 standard; protein; 136 AA.

XX AAW03088;

DT 19-FEB-1997 (first entry)

DE Staphylokinase derivative M21 with altered charge cluster 21.

KW Staphylokinase; mutant; mutein; variant; immunogenicity; decrease;  
derivative; SakSTAR; arterial thrombosis; thrombolytic agent.

OS Staphylococcus aureus.  
Synthetic.

XX Key Location/Qualifiers  
8.10

FT /label= altered\_M21\_cluster  
FT /note= "wild-type cluster 21 sequence Lys-Tyr-Lys  
FT has been replaced by Ala-Tyr-Ala to  
FT reduce immunogenicity"

PN WO9621016-A2.

PD 11-JUL-1996.

PF 03-JAN-1996; 96WO-EP00081.

PR 17-NOV-1995; 95JP-0299781.

PR 06-JAN-1995; 95EP-0200023.

PR 11-JAN-1995; 95US-0371505.

PR 09-JUN-1995; 95EP-0201531.

PR 06-JUL-1995; 95US-0499092.

XX  
PA (COLL/) COLLEN D.  
PA (LEUV-) LEUVEN RES & DEV VZW.  
XX

PI Collen D;

XX  
DR WPI; 1996-333991/33.

XX  
PT New staphylokinase derivs. having reduced immunogenicity - useful  
PT for treating arterial thrombosis

XX  
PS Example 2; Page -; 58pp; English.

XX  
CC Staphylokinase derivatives showing a reduced immunogenicity as  
CC compared to wild-type staphylokinase are claimed. The derivatives  
CC are useful as thrombolytic agents to treat arterial thrombosis and  
CC are pref. produced by eliminating at least one of the epitopes  
CC indicated in the features table. The epitopes are destroyed by  
CC replacing one or more amino acid residues in a charge cluster by an  
CC Ala residue. Mutations are introduced using site-directed  
CC mutagenesis on wild-type staphylokinase genomic DNA from the S.aureus  
CC lysogenic strain SakSTAR. The present sequence is a specific  
CC example of a staphylokinase derivative according to the invention;  
CC the sequence does not appear in the specification but can be generated  
CC from the wild-type staphylokinase sequence using the description given  
CC in Table 1 of the disclosure.

XX  
SQ Sequence 136 AA;

Query Match 98.3%; Score 703; DB 17; Length 136;  
Best Local Similarity 98.5%; Pred. No. 3.9e-69;  
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SSSFDDKGGKKGGDASFEPTGYLMVNTGVDSKGNELLSPHYVEFPKPGTTLTKKEKI 60  
Db 1 sssfdkgayakgddasyfeptgpylmvntgvdsksqnelisphyvefpkpgtlttkexi 60

OY 61 EYVVEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFVVPDLSEHI 120  
Db 61 eyvewaldataykefrveldpsakievttydknkkkeetksfptekegfvpdlsehi 120

OY 121 KNPGFNLITTKVIEKK 136  
Db 121 knpgfnlittkviekk 136

Search completed: September 1, 2002, 03:17:40  
Job time: 7989 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2002, 01:58:01 ; Search time 24.89 Seconds  
(without alignments)  
133.462 Million cell updates/sec

Title: US-09-728-670-10  
Perfect score: 715  
Sequence: 1 SSSFDKGYKKGGDASYFEP.....SEHIKNPGFNLTITKVIEKK 136

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCRTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	136	1 US-08-371-505-2	Sequence 2, Appl1
2	715	100.0	136	1 US-08-784-971-5	Sequence 5, Appl1
3	715	100.0	137	1 US-08-256-261-6	Sequence 6, Appl1
4	715	100.0	137	3 US-08-852-299-6	Sequence 6, Appl1
5	711	99.4	137	1 US-08-256-261-4	Sequence 4, Appl1
6	711	99.4	137	3 US-08-852-299-4	Sequence 4, Appl1
7	705	98.6	137	1 US-08-075-545-1	Sequence 1, Appl1
8	695	97.2	137	1 US-08-256-261-2	Sequence 2, Appl1
9	695	97.2	137	3 US-08-852-299-2	Sequence 2, Appl1
10	695	97.2	163	1 US-08-256-261-17	Sequence 17, Appl1
11	695	97.2	163	3 US-08-852-299-17	Sequence 17, Appl1
12	692	96.8	137	1 US-08-256-261-14	Sequence 14, Appl1
13	692	96.8	137	3 US-08-852-299-14	Sequence 14, Appl1
14	689	96.4	137	1 US-08-256-261-12	Sequence 12, Appl1
15	689	96.4	137	3 US-08-852-299-12	Sequence 12, Appl1
16	653	91.3	126	1 US-08-075-545-2	Sequence 2, Appl1
17	643	89.9	127	1 US-08-256-261-8	Sequence 8, Appl1
18	643	89.9	127	3 US-08-852-299-8	Sequence 8, Appl1
19	620	86.7	123	1 US-08-852-299-10	Sequence 10, Appl1
20	620	86.7	123	3 US-08-852-299-10	Sequence 10, Appl1
21	79	11.0	15	4 US-09-561-108-19	Sequence 19, Appl1
22	79	11.0	15	4 US-09-561-108-19	Sequence 19, Appl1
23	76	10.6	15	4 US-09-561-108-20	Sequence 20, Appl1
24	76	10.6	15	4 US-09-561-108-20	Sequence 20, Appl1
25	74	10.3	592	4 US-09-000-145-6	Sequence 6, Appl1
26	73.5	10.3	2522	4 US-09-251-645-13	Sequence 13, Appl1
27	73	10.2	14	2 US-08-934-222-23	Sequence 23, Appl1

28	73	10.2	14	2 US-08-933-402-23	Sequence 23, Appl1
29	73	10.2	14	2 US-09-207-621-23	Sequence 23, Appl1
30	73	10.2	14	2 US-08-532-818-23	Sequence 23, Appl1
31	73	10.2	14	3 US-09-231-797-23	Sequence 23, Appl1
32	73	10.2	14	3 US-08-934-224-23	Sequence 23, Appl1
33	73	10.2	14	3 US-08-933-843-23	Sequence 23, Appl1
34	73	10.2	14	4 US-08-934-223-23	Sequence 23, Appl1
35	73	10.2	14	4 US-09-413-492-23	Sequence 23, Appl1
36	73	10.2	770	1 US-08-525-654A-1	Sequence 1, Appl1
37	73	10.2	771	1 US-08-525-654A-3	Sequence 3, Appl1
38	73	10.2	2509	1 US-08-469-005A-10	Sequence 10, Appl1
39	73	10.2	2511	4 US-09-261-907-2	Sequence 2, Appl1
40	70.5	9.9	706	1 US-08-339-152A-29	Sequence 29, Appl1
41	70.5	9.9	706	2 US-08-007-999B-4	Sequence 4, Appl1
42	70.5	9.9	706	2 US-08-689-276A-4	Sequence 4, Appl1
43	69	9.7	423	1 US-08-476-008-8	Sequence 8, Appl1
44	69	9.7	423	1 US-08-306-063-8	Sequence 8, Appl1
45	69	9.7	423	1 US-08-833-485-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1  
US-08-371-505-2  
; Sequence 2, Application US/08371505  
; Patent No. 5695754  
; GENERAL INFORMATION:  
; APPLICANT: COLLEN, DESIRE  
; TITLE OF INVENTION: STAPHYLOKINASE DERIVATIVES  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WEBB ZIESENHEIM BREUNING LOGSDON ORKIN & HANSON, P.C.  
; STREET: 700 KOPERS BUILDING, 436 SEVENTH AVENUE  
; CITY: PITTSBURGH  
; STATE: PENNSYLVANIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 15222-2363  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: NEC 286  
; OPERATING SYSTEM: DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; FILING DATE: 11 JAN 1995  
; CLASSIFICATION: 424  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 136  
; TYPE: AMINO ACID  
; STRANDEDNESS: SINGLE  
; TOPOLOGY: UNKNOWN  
; US-08-371-505-2

Query Match 100.0%; Score 715; DB 1; Length 136;  
Best local similarity 100.0%; Pred. No. 2e-78;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSFDKGYKKGGDASYFEP...TGYLMVNTGVDSKGNELLSPHYVEFP...IKPGTTLTKRKI 60  
Db 1 SSSFDKGYKKGGDASYFEP...TGYLMVNTGVDSKGNELLSPHYVEFP...IKPGTTLTKRKI 60  
QY 61 EYVWALDATAFYKEFRVVELDPSAKIEVTYYDKNKKKEETKSP...ITEKGFVVPDLSEHI 120  
Db 61 EYVWALDATAFYKEFRVVELDPSAKIEVTYYDKNKKKEETKSP...ITEKGFVVPDLSEHI 120  
QY 121 KNPGFNLTITKVIEKK 136  
Db 121 KNPGFNLTITKVIEKK 136

## RESULT 2

US-08-784-971-5  
; Sequence 5, Application US/08784971  
; Patent No. 5951980  
; GENERAL INFORMATION:  
; APPLICANT: COLLEN, DESIRE  
; TITLE OF INVENTION: NEW STAPHYLOKINASE DERIVATIVES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THE WEBB LAW FIRM  
; STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE  
; CITY: PITTSBURGH  
; STATE: PENNSYLVANIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 15219-1818  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" FLOPPY DISK  
; COMPUTER: DIGITAL VENTURIS GL 6200  
; OPERATING SYSTEM: DOS  
; SOFTWARE: MICROSOFT WORD 2.0c  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/784,971  
; FILING DATE: 16-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/371,505  
; FILING DATE: 11-JAN-1995  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 136  
; TYPE: AMINO ACID  
; STRANDEDNESS: SINGLE  
; TOPOLOGY: UNKNOWN  
; US-08-784-971-5

## Query Match

Best Local Similarity 100.0%; Score 715; DB 2; Length 136;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSEDKGKYYKKGDASFEPTGPYLMVNTGVDSKGNELLSPHYVEFPKPGTTLTKEKI 60  
DB 1 SSSEDKGKYYKKGDASFEPTGPYLMVNTGVDSKGNELLSPHYVEFPKPGTTLTKEKI 60  
QY 61 EYVWALDATAYKKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
DB 61 EYVWALDATAYKKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
QY 121 KNPGFNLITKVIEKK 136  
DB 121 KNPGFNLITKVIEKK 136

## RESULT 3

US-08-256-261-6  
; Sequence 6, Application US/08256261  
; Patent No. 5801037  
; GENERAL INFORMATION:  
; APPLICANT: Behnke, Detlef  
; APPLICANT: Schlotz, Bernhard  
; APPLICANT: Albrecht, Sybille  
; APPLICANT: G hrs, Karl-Heinz  
; APPLICANT: Hartmann, Manfred  
; TITLE OF INVENTION: Expression of signal-peptide-free  
; TITLE OF INVENTION: staphylokinases  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/256,261  
; FILING DATE:  
; CLASSIFICATION: 435  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 137 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-256-261-6

## Query Match

Best Local Similarity 100.0%; Score 715; DB 1; Length 137;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSEDKGKYYKKGDASFEPTGPYLMVNTGVDSKGNELLSPHYVEFPKPGTTLTKEKI 60  
DB 2 SSSEDKGKYYKKGDASFEPTGPYLMVNTGVDSKGNELLSPHYVEFPKPGTTLTKEKI 61  
QY 61 EYVWALDATAYKKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
DB 62 EYVWALDATAYKKEFRVVELDPSAKIEVYYDKNKKKEETKSPITEKGFVVPDLSEHI 121  
QY 121 KNPGFNLITKVIEKK 136  
DB 122 KNPGFNLITKVIEKK 137

## RESULT 4

US-08-852-299-6  
; Sequence 6, Application US/08852299  
; Patent No. 6010897  
; GENERAL INFORMATION:  
; APPLICANT: Behnke, Detlef  
; APPLICANT: Schlotz, Bernhard  
; APPLICANT: Albrecht, Sybille  
; APPLICANT: G hrs, Karl-Heinz  
; APPLICANT: Hartmann, Manfred  
; TITLE OF INVENTION: Expression of signal-peptide-free  
; TITLE OF INVENTION: staphylokinases  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,299  
; FILING DATE: 17-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/256,261  
; FILING DATE:  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 137 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein



ADDRESSEE: KECK, MAHIN & CATE  
STREET: P.O. BOX 06110  
CITY: CHICAGO  
STATE: ILLINOIS  
COUNTRY: U.S.A.  
ZIP: 60606-0110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS-DOS/PC DOS  
SOFTWARE: DOS Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/075,545  
FILING DATE: 14-JUN-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01722  
FILING DATE: 17-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Fieft, Martin; Gollin, Michael A.  
REGISTRATION NUMBER: 16,900; 31,957  
REFERENCE/DOCKET NUMBER: 47004-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 789-3400  
TELEFAX: (202) 789-1158  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 163 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: 1linear  
MOLECULE TYPE: peptide  
US-08-075-545-1

Query Match 98.6%; Score 705; DB 1; Length 163;  
Best Local Similarity 98.5%; Pred. No. 4.2e-77;  
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 SSSFDKGGKKGGDASFEPTGPLYMNVNTGVDSKGNELLSPHYVEFPKPGTTLTKKEI 60  
DB 28 SSSFDKGGKKGGDASFEPTGPLYMNVNTGVDSKGNELLSPHYVEFPKPGTTLTKKEI 87  
QY 61 EYVEMALDATAKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFPVPLSEHI 120  
DB 88 EYVEMALDATAKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFPVPLSEHI 147  
QY 121 KNPGFNLTQVIEKK 136  
DB 148 KNPGFNLTQVIEKK 163

RESULT 8  
US-08-256-261-2  
Sequence 2, Application US/08256261  
Patent No. 5801037  
GENERAL INFORMATION:  
APPLICANT: Behnke, Detlef  
APPLICANT: Schlott, Bernhard  
APPLICANT: Albrecht, Sybille  
APPLICANT: G hrs, Karl-Heinz  
APPLICANT: Hartmann, Manfred  
TITLE OF INVENTION: Expression of signal-peptide-free  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Neave  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/256,261  
FILING DATE:  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 137 amino acids  
TYPE: amino acid  
TOPOLOGY: 1linear  
MOLECULE TYPE: protein  
US-08-256-261-2

Query Match 97.2%; Score 695; DB 1; Length 137;  
Best Local Similarity 97.8%; Pred. No. 5.1e-76;  
Matches 133; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 SSSFDKGGKKGGDASFEPTGPLYMNVNTGVDSKGNELLSPHYVEFPKPGTTLTKKEI 60  
DB 2 SSSFDKGGKKGGDASFEPTGPLYMNVNTGVDSKGNELLSPHYVEFPKPGTTLTKKEI 61  
QY 61 EYVEMALDATAKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFPVPLSEHI 120  
DB 62 EYVEMALDATAKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFPVPLSEHI 121  
QY 121 KNPGFNLTQVIEKK 136  
DB 122 KNPGFNLTQVIEKK 137

RESULT 9  
US-08-852-299-2  
Sequence 2, Application US/08852299  
Patent No. 6010897  
GENERAL INFORMATION:  
APPLICANT: Behnke, Detlef  
APPLICANT: Schlott, Bernhard  
APPLICANT: Albrecht, Sybille  
APPLICANT: G hrs, Karl-Heinz  
APPLICANT: Hartmann, Manfred  
TITLE OF INVENTION: Expression of signal-peptide-free  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Neave  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/852,299  
FILING DATE: 17-MAY-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/256,261  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 137 amino acids  
TYPE: amino acid  
TOPOLOGY: 1linear  
MOLECULE TYPE: protein  
US-08-852-299-2



Query Match	97.28;	Score 695;	DB 3;	Length 137;
Best Local Similarity	97.88;	Pred. No. 5.1e-76;		
Matches 133; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

```
QY      1 SSSFDKGKXKKGGDDASYFEPTGPYLNVNTGVDSKGNELLSHYVEFPICKP GTTLTKRKI   60
        |||||
Db       2 SSSFDKGKYKKGGDDASYFEPTGPYLNVNTGVDSKRNELLSPRIVEFPIKP GTTLTKRKI   61
QY      61 EYYVEMALDATAAYKEFRVELDPSAKIEVTYYDKNKKKEETKSPITEKG FVPDLSEHT   120
        |||||
Db       62 EYYVEMALDATAAYKEFRVELDPSAKIEVTYYDKNKKKEETKSPITEKG FVPDLSEHT   121
```

Qy	121	KNPGFNLTQVIEKK	136
Db	122	KNPGFNLTQVIEKK	137

RESULT 10  
US-08-256-261-17  
; Sequence 17, Application US/08256261

```

; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlott, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; TITLE OF INVENTION: staphylokinases
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
;

```

Query Match	97.2%;	Score 695;	DB 1;	Length 163;
Best Local Similarity	97.8%;	Pred. No. 6.7e-76;		
Matches 133; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

[illegible]

RESULT 11  
US-08-852-299-17  
; Sequence 17, Application US/08852299  
; Patent No. 6010897

```

; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlott, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; TITLE OF INVENTION: staphylokinases
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
;

```

Query Match	97.28;	Score 695;	DB 3;	Length 163;
Best Local Similarity	97.88;	Pred. No. 6.7e-76;		
Matches 133; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

[illegible]

```

RESULT 12
US-08-256-261-14
; Sequence 14, Application US/08256261
; Patent No. 5801037
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlott, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; TITLE OF INVENTION: staphylokinases
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Neave
;

```

STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/256,261  
FILING DATE:  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 137 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-256-261-14

Query Match 96.8%; Score 692; DB 1; Length 137;  
Best Local Similarity 97.1%; Pred. No. 1.2e-75;  
Matches 132; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 SSSFDKGGKKGGDASFEPTGPLYMNVNTGVDKSGNELSPHYVEFPKPGTTLTKEKI 60  
|||||  
DB 2 SSSFDKGGKKGGDASFEPTGPLYMNVNTGVDKRNELSPRYVEFPKPGTTLTKEKI 61  
OY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFPVPLSEHI 120  
|||||  
DB 62 EYVEMALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFPVPLSEHI 121  
OY 121 KNPGENLITKVIEKK 136  
|||||  
DB 122 KNPGENLITKVIEKK 137

RESULT 13  
US-08-852-299-14  
Sequence 14, Application US/08852299  
Patent No. 6010897  
GENERAL INFORMATION:  
APPLICANT: Behnke, Detlef  
APPLICANT: Schlott, Bernhard  
APPLICANT: Albrecht, Sybille  
APPLICANT: G hrs, Karl-Heinz  
APPLICANT: Hartmann, Manfred  
TITLE OF INVENTION: Expression of signal-peptide-free  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Neave  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/852,299  
FILING DATE: 17-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/256,261  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 137 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-852-299-14

Query Match 96.8%; Score 692; DB 3; Length 137;  
Best Local Similarity 97.1%; Pred. No. 1.2e-75;  
Matches 132; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 SSSFDKGGKKGGDASFEPTGPLYMNVNTGVDKSGNELSPHYVEFPKPGTTLTKEKI 60  
|||||  
DB 2 SSSFDKGGKKGGDASFEPTGPLYMNVNTGVDKRNELSPRYVEFPKPGTTLTKEKI 61  
OY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFPVPLSEHI 120  
|||||  
DB 62 EYVEMALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFPVPLSEHI 121  
OY 121 KNPGENLITKVIEKK 136  
|||||  
DB 122 KNPGENLITKVIEKK 137

RESULT 14  
US-08-256-261-12  
Sequence 12, Application US/08256261  
Patent No. 5801037  
GENERAL INFORMATION:

APPLICANT: Behnke, Detlef  
APPLICANT: Schlott, Bernhard  
APPLICANT: Albrecht, Sybille  
APPLICANT: G hrs, Karl-Heinz  
APPLICANT: Hartmann, Manfred  
TITLE OF INVENTION: Expression of signal-peptide-free  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Neave  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/256,261  
FILING DATE:  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 137 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-256-261-12

Query Match 96.4%; Score 689; DB 1; Length 137;  
Best Local Similarity 97.1%; Pred. No. 2.7e-75;  
Matches 132; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 SSSFDKGGKKGGDASFEPTGPLYMNVNTGVDKSGNELSPHYVEFPKPGTTLTKEKI 60  
|||||  
DB 2 SSSFDKGGKKGGDASFEPTGPLYMNVNTGVDKRNELSPRYVEFPKPGTTLTKEKI 61  
OY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFPITEKGFPVPLSEHI 120  
|||||

Db	62	EYVEMALDATAIKFRVVELDPSAKIEVTYYDDKNKKKEETKSPITEKGVPDLSEHI	121
QY	121	KNPGFNLITKVIEKK	136
Db	122	KNPGFNLITKVIEKK	137

RESULT 15  
US-08-852-299-12  
; Sequence 12, Application US/08852299

```

; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schloft, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; TITLE OF INVENTION: staphylokinases
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
;

```

```

; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Floppy disk
; ; COMPUTER: IBM PC compatible
; ; OPERATING SYSTEM: PC-DOS/MS-DOS
; ; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; ; CURRENT APPLICATION DATA:
; ;

```

Query Match	96.4%;	Score 689;	DB 3;	Length 137;
Best Local Similarity	97.1%;	Pred. No. 2.7e-75;		
Matches 132; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

Oy		61 EYVEMALDATAKKEFRVELDPSAKIEVVYYDKNKKKEETKSFPITEKGVPDLSEHI	120
Db		62 EYVEMALDATAKKEFRVELDPSAKIEVVYYDKNKKKEETKSFPITEKGVPDLSEHI	121

QY	121	KNPGFNLITKVIEKK	136
Db	122	KNPGFNLITKVIEKK	137

RESULT 16  
US-08-075-545-2  
; Sequence 2, Application US/08075545  
; Patent No. 5475089  
; GENERAL INFORMATION:  
; APPLICANT: MATSUO, Osamu; SAKAI, Masashi; SHIMURA,  
; APPLICANT: KISAKU; SANSAWA, Hiroshi; WATANABE,  
; APPLICANT: TSUNEKAZU; MATSUMOTO, Tsuneo; SHISHIDO,

```

1  APPLICANT:  YOSHIYUKI, HASHIMOTO, Shusuke; YOKOKURA
2  APPLICANT:  Teruo; ONOE, Masaharu; SAKO, Tomoyuki
3  TITLE OF INVENTION:  THROMBOLYTIC AGENT
4  NUMBER OF SEQUENCES:  2
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE:  KECK, MAHIN & CATE
7  STREET:  P.O. BOX 06110
8  CITY:  CHICAGO
9  STATE:  ILLINOIS
10 COUNTRY:  U.S.A.
11 ZIP:  60606-0110
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE:  Floppy disk
15 COMPUTER:  IBM PC compatible
16 OPERATING SYSTEM:  MS-DOS/PC DOS
17 SOFTWARE:  DOS Text
18
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER:  US/08/075,545
21 FILING DATE:  14-JUN-1993
22
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER:  PCT/JP91/01722
25 FILING DATE:  17-DEC-1991
26
27 ATTORNEY/AGENT INFORMATION:
28 NAME:  Fleit, Martin; Gollin, Michael A.
29 REGISTRATION NUMBER:  16,900; 31,957
30 REFERENCE/DOCKET NUMBER:  47004-015
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE:  (202) 789-3400
33 TELEFAX:  (202) 789-1158
34
35 INFORMATION FOR SEQ ID NO:  2:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH:  126 amino acids
38 TYPE:  amino acid
39 STRANDEDNESS:  unknown
40 TOPOLOGY:  linear
41
42 MOLECULE TYPE:  peptide
43
44 US-08-075-545-2

```

Query Match	91.38;	Score 653;	DB 1;	length 126;
Best Local Similarity	98.48;	Pred. No. 5.1e-71;		
Matches 124; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Oy		71 TAYKEFRVELDPSAKIEVTTYDDKNKKKKEETKSFPITEKGFWPDLSEHIKNPGFNLTIK	130
Db		61 TAYKEFRVELDPSAKIEVTTYDDKNKKKKEETKSFPITEKGFWPDLSEHIKNPGFNLTIK	120

QY	131	VVIEKK	136
Db	121	VVIEKK	126

RESULT 17  
 US-08-256-261-8  
 ; Sequence 8, Application US/08256261  
 ; Patent No. 5801037  
 ;  
 ; GENERAL INFORMATION:  
 ;  
 ; APPLICANT: Behnke, Detlef  
 ;  
 ; APPLICANT: Schlott, Bernhard  
 ;  
 ; APPLICANT: Albrecht, Sybille  
 ;  
 ; APPLICANT: G hrs, Karl-Heinz  
 ;  
 ; APPLICANT: Hartmann, Manfred  
 ;  
 ; TITLE OF INVENTION: Expression of signal-peptide-free  
 ;  
 ; TITLE OF INVENTION: staphylokinases  
 ;  
 ; NUMBER OF SEQUENCES: 40  
 ;  
 ; CORRESPONDENCE ADDRESS:  
 ;  
 ; ADDRESSEE: Fish & Neave  
 ;  
 ; STREET: 1251 Avenue of the Americas  
 ;  
 ; CITY: New York  
 ;

STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/256,261  
FILING DATE:  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 127 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-256-261-8

Query Match 89.9%; Score 643; DB 1; Length 127;  
Best Local Similarity 97.6%; Pred. No. 8.3e-70;  
Matches 123; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 11 KGDDASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKIEYVEMALDA 70  
|||||  
Db 2 KGDDASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKIEYVEMALDA 61

QY 71 TAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHIKNPGFNLTITK 130  
|||||  
Db 62 TAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHIKNPGFNLTITK 121

QY 131 VVIEKK 136  
|||||  
Db 122 VVIEKK 127

RESULT 18  
US-08-852-299-8

Sequence 8, Application US/088522299  
Patent No. 6010897

GENERAL INFORMATION:

APPLICANT: Behnke, Detlef

APPLICANT: Schlott, Bernhard

APPLICANT: Albrecht, Sybille

APPLICANT: G hrs, Karl-Heinz

APPLICANT: Hartmann, Manfred

TITLE OF INVENTION: Expression of signal-peptide-free

TITLE OF INVENTION: staphylokinases

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Neave

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/852,299

FILING DATE: 17-MAY-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/256,261

FILING DATE:

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 127 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-852-299-8

Query Match 89.9%; Score 643; DB 3; Length 127;  
Best Local Similarity 97.6%; Pred. No. 8.3e-70;  
Matches 123; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 11 KGDDASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKIEYVEMALDA 70  
|||||  
Db 2 KGDDASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKIEYVEMALDA 61

QY 71 TAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHIKNPGFNLTITK 130  
|||||  
Db 62 TAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHIKNPGFNLTITK 121

QY 131 VVIEKK 136  
|||||  
Db 122 VVIEKK 127

RESULT 19  
US-08-256-261-10

Sequence 10, Application US/08256261  
Patent No. 5801037

GENERAL INFORMATION:

APPLICANT: Behnke, Detlef

APPLICANT: Schlott, Bernhard

APPLICANT: Albrecht, Sybille

APPLICANT: G hrs, Karl-Heinz

APPLICANT: Hartmann, Manfred

TITLE OF INVENTION: Expression of signal-peptide-free

TITLE OF INVENTION: staphylokinases

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Neave

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/256,261

FILING DATE:

CLASSIFICATION: 435

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 123 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-256-261-10

Query Match 86.7%; Score 620; DB 1; Length 123;  
Best Local Similarity 97.5%; Pred. No. 4.6e-67;  
Matches 119; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 15 ASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKIEYVEMALDATAK 74  
|||||  
Db 2 ASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKIEYVEMALDATAK 61

QY 75 EFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHIKNPGFNLTITK 134  
|||||  
Db 62 EFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHIKNPGFNLTITK 121





; SEQ ID NO 20  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-561-500-20

Query Match 10.6%; Score 76; DB 4; Length 15;  
Best Local Similarity 93.3%; Pred. No. 0.0057;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSFDKGYKKGGDDA 15  
|||  
Db 1 SSSFDKGYKKGGDDA 15

RESULT 24  
US-09-561-108-20

; Sequence 20, Application US/09561108  
; Patent No. 6342221

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Rolf A. Breken

; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF

; FILE REFERENCE: 4001.002584

; CURRENT APPLICATION NUMBER: US/09/561,108

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 1999-04-28

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 20

; LENGTH: 15

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

; OTHER INFORMATION: PEPTIDE

Query Match 10.6%; Score 76; DB 4; Length 15;  
Best Local Similarity 93.3%; Pred. No. 0.0057;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSSFDKGYKKGGDDA 15  
|||  
Db 1 SSSFDKGYKKGGDDA 15

RESULT 25

US-09-000-145-6

; Sequence 6, Application US/09000145

; Patent No. 6169172

; GENERAL INFORMATION:

; APPLICANT: DEVAUCHELLE, Gerrard

; APPLICANT: GARNIER, Laurence

; APPLICANT: CAHOREAU, Claire

; APPLICANT: CERUTTI, Martine

; TITLE OF INVENTION: USE OF A PROLACTIN RECEPTOR OR GROWTH HORMONE RECEPTOR

; FILE REFERENCE: 0384-0047-0XPCT

; CURRENT APPLICATION NUMBER: US/09/000,145

; EARLIER FILING DATE: 1998-03-16

; EARLIER FILING DATE: 1996-08-02

; EARLIER FILING DATE: 1995-08-02

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 6  
; LENGTH: 592  
; TYPE: PRT  
; ORGANISM: Rattus rattus  
US-09-000-145-6

Query Match 10.3%; Score 74; DB 4; Length 592;  
Best Local Similarity 23.9%; Pred. No. 2.3;  
Matches 34; Conservative 17; Mismatches 43; Indels 48; Gaps 7;

QY 4 FDKGYKKGGDDASY-----FEPTGPY--LMVNTGVDSKGNLSP-HYVEFP---I 49  
|||  
Db 256 FDTLLEKKGSEELISALGCQDFPPTSDCEDLVEFLEVDNEDERLMPHSKEYPGQGV 315

QY 50 KPGTTLTKKEIEYVEMALDATAKFERVELDPSAKIEVYYDK---NKKKEETKSP 105  
|||  
Db 316 KP-----THLPPDSGSHGSHSLSEKCEEPQAYP 348

QY 106 ITEKGFVVPDLSEHKNPFNL 127  
|  
Db 349 PT--LHIEPTEKPEPNEANI 367

RESULT 26

US-09-251-645-13

; Sequence 13, Application US/09251645

; Patent No. 6281413

; GENERAL INFORMATION:

; APPLICANT: Kramer, Vance C.

; APPLICANT: Morgan, Michael K.

; APPLICANT: Anderson, Arne R.

; APPLICANT: Hart, Hope

; APPLICANT: Warren, Gregory W.

; APPLICANT: Dunn, Martha

; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS

; FILE REFERENCE: CGC1963/A

; CURRENT APPLICATION NUMBER: US/09/251,645

; CURRENT FILING DATE: 1999-02-17

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 13

; LENGTH: 2522

; TYPE: PRT

; ORGANISM: Photorhabdus luminescens

Query Match 10.3%; Score 73.5; DB 4; Length 2522;  
Best Local Similarity 28.9%; Pred. No. 23;  
Matches 37; Conservative 16; Mismatches 42; Indels 33; Gaps 8;

QY 12 GDD---ASYFEPTGPYLMVNTGVDSKG-----NELSPHYVEFPKPGT---T 54  
|||  
Db 1461 GDDYATDSLKNPNDLKQYYMT--DSKGTATDVSGPYDINTAISPAAVQYVYKAGSKEQT 1518

QY 55 LTKK-IEYVEMALDATAKFERVELD-----PSAKIEVY---YDKNKKKEET 101  
|  
Db 1519 FTADKNVSIQSPSPDENMY-QFNALIDGSSLNFTNNSASIDITFTAFADGRRIGYES 1577

QY 102 KSPITEK 109  
|  
Db 1578 FSIPITRK 1585

RESULT 27

US-08-934-222-23

; Sequence 23, Application US/08934222

; Patent No. 5928896

; GENERAL INFORMATION:

; APPLICANT: EVANS, Herbert J.

```

APPLICANT: KINI, R. Manjunatha
TITLE OF INVENTION: Polypeptides That Include Conformation-
TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction
TITLE OF INVENTION: Site
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: Suite 500, 3000 K Street NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20007
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,222
FILING DATE: 19-SEPT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/532,818
FILING DATE: 03-MAY-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/143,364
FILING DATE: 29-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/051,741
FILING DATE: 23-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Isacson, John P.
REGISTRATION NUMBER: 33,751
REFERENCE/DOCKET NUMBER: 040433/0148
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-934-222-23

Query Match      10.2%; Score 73; DB 2; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.012;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      11 KGDDASYFEPTGPY 24
        1 ||||||||||||
Db       1 KPDDASYFEPTGPY 14

RESULT 28
US-08-933-402-23
: Sequence 23, Application US/08933402
: Patent No. 5948887
: GENERAL INFORMATION:
: APPLICANT: EVANS, Herbert J.
: APPLICANT: KINI, R. Manjunatha
: TITLE OF INVENTION: Polypeptides That Include Conformation-
: TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction
: TITLE OF INVENTION: Site
: NUMBER OF SEQUENCES: 153
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: Suite 500, 3000 K Street NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20007
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/933,402
: FILING DATE: 19-SEPT-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/532,818
: FILING DATE: 03-MAY-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: U.S. 08/143,364
: FILING DATE: 29-OCT-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: U.S. 08/051,741
: FILING DATE: 23-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Isacson, John P.
: REGISTRATION NUMBER: 33,751
: REFERENCE/DOCKET NUMBER: 040433/0148
: INFORMATION FOR SEQ ID NO: 23:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 14 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
:
: US-08-933-402-23
:
Query Match 10.2%; Score 73; DB 2; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.012;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 11 KGDDASYFEPTGPY 24
| | | | | | | | | | | | | | | |
Db 1 KPDDASYFEPTGPY 14
:
RESULT 29
: US-09-207-621-23
: Sequence 23, Application US/09207621
: Patent No. 5952465
: GENERAL INFORMATION:
: APPLICANT: EVANS, Herbert J.
: APPLICANT: KINI, R. Manjunatha
: TITLE OF INVENTION: Polypeptides That Include Conformation-
: TITLE OF INVENTION: Constraining Groups Which Flank A Proten-Protein Interactio
: NUMBER OF SEQUENCES: 153
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: Suite 500, 3000 K Street NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20007
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/207,621
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/532,818
: FILING DATE: 03-MAY-1996
: APPLICATION NUMBER: PCT/US94/04294
: FILING DATE: 21-APR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: U.S. 08/143,364
: FILING DATE: 29-OCT-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: U.S. 08/051,741
: FILING DATE: 23-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Isacson, John P.
: REGISTRATION NUMBER: 33,751
: REFERENCE/DOCKET NUMBER: 040433/0148

```

;; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14 amino acids  
; TYPE: amino acid  
; TOPOLOGY: 1linear  
US-09-207-621-23

Query Match 10.2%; Score 73; DB 2; Length 14;  
Best Local Similarity 92.9%; Pred. No. 0.012;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 11 KGDDASYFEPTGPY 24  
| | | | | | | | | |  
Db 1 KPDDASYFEPTGPY 14

RESULT 30  
US-08-532-818-23

; Sequence 23, Application US/08532818  
; Patent No. 5965698

; GENERAL INFORMATION:

; APPLICANT: EVANS, Herbert J.

; APPLICANT: KINI, R. Manjunatha

; TITLE OF INVENTION: Polypeptides That Include Conformation-

; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction

; TITLE OF INVENTION: Site

; NUMBER OF SEQUENCES: 153

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: Suite 500, 3000 K Street NW

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20007

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/532,818

; FILING DATE: 03-MAY-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/04294

; FILING DATE: 21-APR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: U.S. 08/143,364

; FILING DATE: 29-OCT-1993

; APPLICATION DATA:

; APPLICATION NUMBER: U.S. 08/051,741

; FILING DATE: 23-APR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Isacson, John P.

; REGISTRATION NUMBER: 33,751

; REFERENCE/DOCKET NUMBER: 040433/0148

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 14 amino acids

; TYPE: amino acid

; TOPOLOGY: 1linear

US-08-532-818-23

Query Match 10.2%; Score 73; DB 2; Length 14;  
Best Local Similarity 92.9%; Pred. No. 0.012;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 11 KGDDASYFEPTGPY 24  
| | | | | | | | | |  
Db 1 KPDDASYFEPTGPY 14

RESULT 31  
US-09-231-797-23

; Sequence 23, Application US/09231797  
; Patent No. 6084066

; GENERAL INFORMATION:

; APPLICANT: EVANS, Herbert J.

; APPLICANT: KINI, R. Manjunatha

; TITLE OF INVENTION: Polypeptides That Include Conformation-

; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction

; TITLE OF INVENTION: Site

; NUMBER OF SEQUENCES: 153

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: Suite 500, 3000 K Street NW

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20007

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/231,797

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/532,818

; FILING DATE: 03-MAY-1996

; APPLICATION NUMBER: PCT/US94/04294

; FILING DATE: 21-APR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: U.S. 08/143,364

; FILING DATE: 29-OCT-1993

; APPLICATION DATA:

; APPLICATION NUMBER: U.S. 08/051,741

; FILING DATE: 23-APR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Isacson, John P.

; REGISTRATION NUMBER: 33,751

; REFERENCE/DOCKET NUMBER: 040433/0148

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 14 amino acids

; TYPE: amino acid

; TOPOLOGY: 1linear

US-09-231-797-23

Query Match 10.2%; Score 73; DB 3; Length 14;  
Best Local Similarity 92.9%; Pred. No. 0.012;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 11 KGDDASYFEPTGPY 24  
| | | | | | | | | |  
Db 1 KPDDASYFEPTGPY 14

RESULT 32  
US-08-934-224-23

; Sequence 23, Application US/08934224  
; Patent No. 6100044

; GENERAL INFORMATION:

; APPLICANT: EVANS, Herbert J.

; APPLICANT: KINI, R. Manjunatha

; TITLE OF INVENTION: Polypeptides That Include Conformation-

; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction

; TITLE OF INVENTION: Site

; NUMBER OF SEQUENCES: 153

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: Suite 500, 3000 K Street NW

; CITY: Washington

; STATE: DC



COUNTRY: USA  
ZIP: 20007  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/934,224  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/532,818  
FILING DATE: 03-MAY-1996  
APPLICATION NUMBER: PCT/US94/04294  
FILING DATE: 21-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/143,364  
FILING DATE: 29-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/051,741  
FILING DATE: 23-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Isacson, John P.  
REGISTRATION NUMBER: 33,751  
REFERENCE/DOCKET NUMBER: 040433/0148  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-934-224-23

Query Match 10.2%; Score 73; DB 3; Length 14;  
Best Local Similarity 92.9%; Pred. No. 0.012;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 KDDASYEPTGPY 24  
Db 1 KPDDASYEPTGPY 14

RESULT 33  
US-08-933-843-23  
Sequence 23, Application US/08933843  
Patent No. 611069  
GENERAL INFORMATION:  
APPLICANT: EVANS, Herbert J.  
APPLICANT: KINI, R. Manjunatha  
TITLE OF INVENTION: Polypeptides That Include Conformation-  
TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction  
TITLE OF INVENTION: Site  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: Suite 500, 3000 K Street NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20007  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/933,843  
FILING DATE: 19-SEPT-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/532,818  
FILING DATE: 03-MAY-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/143,364

FILING DATE: 29-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/051,741  
FILING DATE: 23-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Isacson, John P.  
REGISTRATION NUMBER: 33,751  
REFERENCE/DOCKET NUMBER: 040433/0148  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-933-843-23

Query Match 10.2%; Score 73; DB 3; Length 14;  
Best Local Similarity 92.9%; Pred. No. 0.012;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 KDDASYEPTGPY 24  
Db 1 KPDDASYEPTGPY 14

RESULT 34  
US-08-934-223-23  
Sequence 23, Application US/08934223  
Patent No. 6147189  
GENERAL INFORMATION:  
APPLICANT: EVANS, Herbert J.  
APPLICANT: KINI, R. Manjunatha  
TITLE OF INVENTION: Polypeptides That Include Conformation-  
TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interacti  
TITLE OF INVENTION: Site  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: Suite 500, 3000 K Street NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20007  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/934,223  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/532,818  
FILING DATE: 03-MAY-1996  
APPLICATION NUMBER: PCT/US94/04294  
FILING DATE: 21-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/143,364  
FILING DATE: 29-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/051,741  
FILING DATE: 23-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Isacson, John P.  
REGISTRATION NUMBER: 33,751  
REFERENCE/DOCKET NUMBER: 040433/0148  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-934-223-23



```
; TITLE OF INVENTION: TRANSGLUTAMINASE ORIGINATED FROM
; TITLE OF INVENTION: CRASSOSTREA GIGAS
; NUMBER OF SEQUENCES: 150
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,654A
; FILING DATE: 28-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6/8283
; FILING DATE: 28-JAN-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 7/3876
; FILING DATE: 13-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-760-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 771 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Crassostrea gigas
; US-08-525-654A-3

Query Match 10.2%; Score 73; DB 1; Length 771;
Best Local Similarity 25.3%; Pred. No. 4.5;
Matches 21; Conservative 22; Mismatches 30; Indels 10; Gaps 4;

QY 25 LNVNVTGVDSKGNELSPHYVEF--PIKPG----TTLTKETIEYVWALDATA-YKEF 76
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 689 LVLTAQAEIAYEGSGLAPQTFIDISSPIKPGDEVKKTIVLRPKRPYY--WGRELIAFTFSK 746
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 77 RVVELDPSAKIEVTTYDKNKKKE 99
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 747 QIVDIETSDADIKVIRONKNDSD 769

RESULT 38
US-08-469-005A-10
; Sequence 10, Application US/08469005A
; Patent No. 5665874
; GENERAL INFORMATION:
; APPLICANT: KUHAJDA, FRANCIS P.
; APPLICANT: PASTERNAK, GARY A.
; TITLE OF INVENTION: CANCER RELATED ANTIGEN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BAKER & BOTTS, L.L.P.
; STREET: 1299 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
```

```
; ZIP: 20004-2400
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,005A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/188,426
; FILING DATE: 24-JAN-1994
; APPLICATION NUMBER: 08/096,908
; FILING DATE: 26-JUL-1993
; APPLICATION NUMBER: 07/917,716
; FILING DATE: 24-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Posorske, Laurence H
; REGISTRATION NUMBER: 34,698
; REFERENCE/DOCKET NUMBER: 062482-0113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-639-7700
; TELEFAX: 202-639-7890
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2509 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-469-005A-10

Query Match 10.2%; Score 73; DB 1; Length 2509;
Best Local Similarity 34.4%; Pred. No. 26;
Matches 21; Conservative 10; Mismatches 28; Indels 2; Gaps 1;

QY 27 VNVTVGVDSKGNELSPHYVEFPIKPGTTLTKETIEYVWALDATA-YKEFRVELDPSAK 86
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 803 LHLSGIDANPNALFPP--VEFPAPRGTPILISPLIKWDHSLAWDVPAEDFPNGSGSPSAA 860
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 87 I 87
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 861 I 861

RESULT 39
US-09-261-907-2
; Sequence 2, Application US/09261907A
; Patent No. 6294364
; GENERAL INFORMATION:
; APPLICANT: ELLIS, CATHERINE
; APPLICANT: LONSDALE, JOHN
; APPLICANT: BERGSMAN, DEK J.
; APPLICANT: MOONEY, JEFFREY L.
; APPLICANT: DEPIERA, MEGAN E.
; APPLICANT: CHAPMAN, CONRAD
; TITLE OF INVENTION: HUMAN FAS
; FILE REFERENCE: GP-70603
; CURRENT APPLICATION NUMBER: US/09/261,907A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2511
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
```





```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Ave., NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/689,276A
FILING DATE: 06-AUG-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/007,999
FILING DATE: 21-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,642
FILING DATE: 20-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,022
FILING DATE: 17-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: JORGE A. GOLDSTEIN
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 0609.3520003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2571
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-689-276A-4

Query Match          9.9%; Score 70.5; DB 2; Length 706;
Best Local Similarity 23.3%; Pred. No. 8;
Matches 27; Conservative 18; Mismatches 56; Indels 15; Gaps 3;

QY 11 KGDDASYEPTGPTGLMNVNTGVDKSGNELLSPHYVEFPPIKPGTTLTKEKIEYVEMALDA 70
   ||||: ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 280 KGDDYNENPTPE-----GSDGTMSDKETHDVKVPTP---LPTNDVDVYFETSDADD 329
   ||||: ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 71 TAYKERVVELDPSAKIEVYYDKNKKKEE-----TKSFPTTEKGFVVPDLSEHIK 121
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 330 NEHARFOKAKEQLIERHRNMDRVKKKEWEAEALQAKNLPKAEKOTLIQHQAQAVK 385
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 43
US-08-476-008-8
; Sequence 8, Application US/08476008
; Patent No. 5627061
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Padgett, Stephen R.
; APPLICANT: Stallings, William C.
; TITLE OF INVENTION: Glycosate Tolerant
; TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F
; STREET: 700 Chesterfield Village Parkway
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,008
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/306,063
; FILING DATE: 13-SEP-1994
; APPLICATION NUMBER: US 07/749,611
; FILING DATE: 28-AUG-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/576,537
; FILING DATE: 31-AUG-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoerner Jr., Dennis R.
; REGISTRATION NUMBER: 30,914
; REFERENCE/DOCKET NUMBER: 38-21(10660)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)537-6099
; TELEFAX: (314)537-6047
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-476-008-8

```

```

Query Match          9.78; Score 69; DB 1; Length 423;
Best Local Similarity 24.58; Pred. No. 5.7;
Matches 26; Conservative 18; Mismatches 40; Indels 22; Gaps 4;

QY 11 KGDASYPEPTGPYIMVNTGVDSKGNELLSPHYVEFPPIK-----PGTTLTKKEIEYY 63
    ||| |  ||: : : : | |: : ||: | ||| |
Db 190 KGDLYS-----KPYIDITLNMKTFGEVIEHQHYQQFVYKGGQSYQSPGT-----YL 236
                                     :||| :| : : : ||

QY 64 VEMALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEK 109
    || ||: |  :||| :| : : : ||
Db 237 VEG--DASSASYFLAAAIKGGTVKVTGIGRNSMGDIRFADYLEK 280
                                     :||| :| : : : ||

RESULT 44
US-08-306-063-8
; Sequence 8, Application US/08306063
; Patent No. 5633435
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Padgett, Stephen R.
; APPLICANT: Stallings, William C.
; TITLE OF INVENTION: Glyphosate Tolerant
; TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BBAF
; STREET: 700 Chesterfield Village Parkway
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;

```

APPLICATION NUMBER: US/08/306,063  
FILING DATE: 13-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/749,611  
FILING DATE: 28-AUG-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/576,537  
FILING DATE: 31-AUG-1990  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoerner Jr., Dennis R.  
REGISTRATION NUMBER: 30,914  
REFERENCE/DOCKET NUMBER: 38-21(10660)A  
TELEPHONE: (314)537-6099  
TELEFAX: (314)537-6047  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 423 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-306-063-8

Query Match 9.7%; Score 69; DB 1; Length 423;  
Best Local Similarity 24.5%; Pred. No. 5.7;  
Matches 26; Conservative 18; Mismatches 40; Indels 22; Gaps 4;

QY 11 KGDDASYEPTGPLYMNVNTGVDSKGNELSPHYVEFPK-----PGTTLTKKEIEY 63  
Db 190 KGDLVS-----KPYIDITLNLMTFGEIEIENQHYQOFVVKGGQSYQSPGT-----YL 236  
QY 64 VEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEK 109  
Db 237 VEG--DASSASYFLAAAIKGTGVTGIGRNSMGDIRFADVLEK 280

## RESULT 45

US-08-833-485-8  
Sequence 8, Application US/08833485  
Patent No. 5804425  
GENERAL INFORMATION:  
APPLICANT: Barry, Gerard F.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Padgette, Stephen R.  
APPLICANT: Stallings, William C.  
TITLE OF INVENTION: Glyphosate Tolerant  
TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F  
STREET: 700 Chesterfield Village Parkway  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/833,485  
FILING DATE: 07-APR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/306,063  
FILING DATE: 13-SEP-1994  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/749,611  
FILING DATE: 28-AUG-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/576,537  
FILING DATE: 31-AUG-1990  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoerner Jr., Dennis R.  
REGISTRATION NUMBER: 30,914  
REFERENCE/DOCKET NUMBER: 38-21(15117)A  
TELEPHONE: (314)737-6099  
TELEFAX: (314)737-6047  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 423 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-833-485-8

Query Match 9.7%; Score 69; DB 1; Length 423;  
Best Local Similarity 24.5%; Pred. No. 5.7;  
Matches 26; Conservative 18; Mismatches 40; Indels 22; Gaps 4;

QY 11 KGDDASYEPTGPLYMNVNTGVDSKGNELSPHYVEFPK-----PGTTLTKKEIEY 63  
Db 190 KGDLVS-----KPYIDITLNLMTFGEIEIENQHYQOFVVKGGQSYQSPGT-----YL 236  
QY 64 VEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEK 109  
Db 237 VEG--DASSASYFLAAAIKGTGVTGIGRNSMGDIRFADVLEK 280

## RESULT 46

US-09-137-440-8  
Sequence 8, Application US/09137440  
Patent No. 6248876  
GENERAL INFORMATION:  
APPLICANT: Barry, Gerard F.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Padgette, Stephen R.  
APPLICANT: Stallings, William C.  
TITLE OF INVENTION: Glyphosate Tolerant  
TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F  
STREET: 700 Chesterfield Village Parkway  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/137,440  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/833,485  
FILING DATE: 07-APR-1997  
APPLICATION NUMBER: US 08/306,063  
FILING DATE: 13-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/749,611

TELEPHONE: (314)537-6047  
TELEFAX: (314)537-6047

US-09-243-374-7

```

:SEQ ID NO:7:
:LENGTH: 427

```





---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2002, 03:18:49 ; Search time 25.12 Seconds  
(without alignments)  
209.628 Million cell updates/sec

Title: US-09-728-670-10  
Perfect score: 715  
Sequence: 1 SSSFDKGRKRGDDASYFEP.....SEHIKNPGFNLTIVIEKK 136

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71.1	99.4	163	SAK_STAM	P00802 staphylococ
2	69.5	97.2	163	SAK_BPP42	P15240 bacterioph
3	80.5	11.3	3097	CADN_DROME	O15943 drosophila
4	78.5	11.0	324	YF02_METJA	O58897 methanococ
5	78.5	11.0	719	IF39_TOBAC	P56821 nicotiana t
6	77.5	10.8	442	TOP5_BPT4	P07065 bacterioph
7	77	10.8	608	PRLR_MOUSE	O08501 mus musculu
8	76.5	10.7	506	VLI_BPV4	P08341 bovine papi
9	75.5	10.6	327	LXB1_PHOLU	P19840 photorhabdu
10	75	10.5	576	NAE2_THEMA	O9x0y0 thermotoga
11	74.5	10.4	2511	FAS_CHICK	P12276 gallus gall
12	74	10.3	298	MOAA_METJA	O58234 methanococ
13	74	10.3	428	V234_FOWPV	P14368 fowlpox vir
14	74	10.3	610	PRLR_RAT	P05710 rattus norv
15	73.5	10.3	430	SERC_ARATH	O96255 arabidopsis
16	73	10.2	231	CTRA_CAUCR	O45994 caulobacter
17	73	10.2	237	TRMD_BUCAL	P57476 buchnera ap
18	73	10.2	263	YFJB_ECOLI	P76612 escherichia
19	73	10.2	2505	FAS_RAT	P12785 rattus norv
20	73	10.2	4969	RYR2_RABIT	P30957 oryctolagus
21	72.5	10.1	374	YEAW_ECOLI	P76253 escherichia
22	72.5	10.1	430	SURA_BUCAL	P57240 buchnera ap
23	72.5	10.1	935	IF38_MEDTR	O9xhml medicago tr
24	72.5	10.1	1043	SYI_CHLPN	O92972 chlamydia p
25	72	10.1	354	YQ30_BACAN	O9rn02 bacillus an
26	72	10.1	429	HISX_MOUSE	O58851 methanococ
27	72	10.1	695	APB2_MOUSE	O06335 mus musculu
28	72	10.1	978	KFMS_RAT	O00495 rattus norv
29	72	10.1	1102	YE20_METJA	O58815 methanococ
30	72	10.1	2663	CENE_HUMAN	O02224 homo sapien
31	71.5	10.0	553	YF61_MUTJA	O58956 methanococ
32	71.5	10.0	560	POTA_MYCPN	P75059 mycoplasma
33	71.5	10.0	667	Y366_MYCGE	P47606 mycoplasma

34	71.5	10.0	686	1	FREL_YEAST	P32791 saccharomyc
35	71	9.9	344	1	CYSA_SYNP7	P14788 synechococc
36	71	9.9	457	1	CRQ_DROME	O27367 drosophila
37	71	9.9	479	1	RFBM_SALTY	P26404 salmonella
38	71	9.9	512	1	PUR9_BACSU	P12048 b bifunctio
39	71	9.9	976	1	KFMS_MOUSE	P09581 mus musculu
40	71	9.9	1021	1	YPT7_CAEEL	P41885 caenorhabdi
41	70.5	9.9	534	1	TCPG_YEAST	P39077 saccharomyc
42	70.5	9.9	559	1	DNL1_PYRKO	O9hnc4 pyrococcus
43	70.5	9.9	608	1	YD56_YEAST	O04399 saccharomyc
44	70	9.8	115	1	YD46_MYCPN	P75432 mycoplasma
45	70	9.8	251	1	Y255_METJA	O57703 methanococ

ALIGNMENTS

RESULT	ID	SAK_STAM	STANDARD	PRT	163 AA.
AC	P00802;	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Staphylokinase precursor (Neutral proteinase) (Protease III).				
GN	SAK OR SAV1944.				
OS	Staphylococcus aureus (strain Mu50 / ATCC 700699), and				
OS	Staphylococcus aureus.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Staphylococcus.				
OX	NCBI_TaxID=158878, 1280;				
RN	[1]	SEQUENCE FROM N.A.			
RP	STRAIN=Mu50 / ATCC 700699;				
RC	MEDLINE=21311952; PubMed=11418146;				
RX	Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,				
RA	Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,				
RA	Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,				
RA	Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,				
RA	Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yabuzaki J.,				
RA	Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,				
RA	Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;				
RT	"Whole genome sequencing of methicillin-resistant Staphylococcus				
RT	aureus.";				
RL	Lancet 357:1225-1240(2001).				
RN	[2]	SEQUENCE FROM N.A.			
RP	MEDLINE=84069795; PubMed=6359061;				
RX	Sako T., Tsuchida N.;				
RA	"Nucleotide sequence of the staphylokinase gene from Staphylococcus				
RT	aureus.";				
RL	Nucleic Acids Res. 11:7679-7693(1983).				
RN	[3]	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 43-163.			
RP	MEDLINE=97290447; PubMed=9145104;				
RX	Radjins A., de Bondt H.L., de Rantier C.;				
RA	"Three-dimensional structure of staphylokinase, a plasminogen				
RT	activator with therapeutic potential.";				
RL	Nat. Struct. Biol. 4:357-360(1997).				
RN	[4]	STRUCTURE BY NMR OF 28-163.			
RP	MEDLINE=98367505; PubMed=9692953;				
RX	Ohlenschlaeger O., Ramchandran R., Guehrs K.H., Schlott B.,				
RA	Brown L.R.;				
RT	"Nuclear magnetic resonance solution structure of the plasminogen-				
RT	activator protein staphylokinase.";				
RL	Biochemistry 37:10635-10642(1998).				
CC	-1- FUNCTION: THIS EXTRACELLULAR PROTEIN IS ONE OF THE PLASMINOGEN				
CC	ACTIVATORS THAT CONVERTS PLASMINOGEN INTO PLASMIN; IT REQUIRES				
CC	CALCIUM ION FOR STABILIZATION.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AP003364; BAB58106.1; -.  
DR EMBL; X00127; CAA24957.1; -.  
DR PIR; A00995; PRSAK.  
DR PDB; 2SAK; 25-FEB-98.  
DR PDB; 1SSN; 02-DEC-98.  
DR InterPro; IPR004093; Staphylokinase.  
DR Pfam; PF02821; Staphylokinase; 1.  
KW Hydrolase; Calcium; Plasminogen activation; signal; 3D-structure;  
KW Complete proteome.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 163 STAPHYLOKINASE.  
SQ SEQUENCE 163 AA; 18490 MW; E56D9FF50AEDE141 CRC64;  
  
Query Match 99.4%; Score 711; DB 1; Length 163;  
Best Local Similarity 99.3%; Pred. No. 4.1e-59;  
Matches 135; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 SSSFDKGYKKGGDASFEPTGPLYMNVNTGVDSKGNELLSPHYVEFPKPGTTLTKEKI 60  
DB 28 SSSFDKGYKKGGDASFEPTGPLYMNVNTGVDSKGNELLSPHYVEFPKPGTTLTKEKI 87  
  
QY 61 EYVEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
DB 88 EYVEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 147  
  
QY 121 KNPGFNLTQVIEKK 136  
DB 148 KNPGFNLTQVIEKK 163  
  
RESULT 2  
SAK\_BPP42  
ID SAK\_BPP42 STANDARD; PRT; 163 AA.  
AC P15240;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Staphylokinase precursor (Neutral proteinase) (Protease III).  
GN SAK.  
OS Bacteriophage P42D.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
OC Lambda phage group.  
OX NCBI\_TaxID=10715;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88121731; PubMed=3123893;  
RA Behnke D., Gerlach D.;  
RT "Cloning and expression in Escherichia coli, Bacillus subtilis, and  
RT Streptococcus sanguis of a gene for staphylokinase -- a bacterial  
RT Plasminogen activator."  
RL Mol. Genet. 210:528-534(1987).  
CC -1- FUNCTION: THIS EXTRACELLULAR PROTEIN IS ONE OF THE PLASMINOGEN  
CC ACTIVATORS THAT CONVERTS PLASMINOGEN INTO PLASMIN; IT REQUIRES  
CC CALCIUM ION FOR STABILIZATION.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X06603; CAA29822.1; -.  
DR EMBL; M57455; AAA98206.1; -.  
DR EMBL; A17537; CAA01341.1; -.  
.

DR PIR; S02330; S02330.  
DR HSSP; P00802; 1SSN.  
DR InterPro; IPR004093; Staphylokinase.  
DR Pfam; PF02821; Staphylokinase; 1.  
KW Hydrolase; Calcium; Plasminogen activation; signal.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 163 STAPHYLOKINASE.  
SQ SEQUENCE 163 AA; 18608 MW; AA2674AF75D36EC CRC64;  
  
Query Match 97.2%; Score 695; DB 1; Length 163;  
Best Local Similarity 97.8%; Pred. No. 1.2e-57;  
Matches 133; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 SSSFDKGYKKGGDASFEPTGPLYMNVNTGVDSKGNELLSPHYVEFPKPGTTLTKEKI 60  
DB 28 SSSFDKGYKKGGDASFEPTGPLYMNVNTGVDSKGNELLSPHYVEFPKPGTTLTKEKI 87  
  
QY 61 EYVEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 120  
DB 88 EYVEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLSEHI 147  
  
QY 121 KNPGFNLTQVIEKK 136  
DB 148 KNPGFNLTQVIEKK 163  
  
RESULT 3  
CADN\_DROME  
ID CADN\_DROME STANDARD; PRT; 3097 AA.  
AC 015943; Q9VJB7;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Neural-cadherin precursor (Cadherin-N protein) (DN-cadherin).  
GN CADN OR CG7100.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Head, and Embryo;  
RX MEDLINE=97388431; PubMed=9247265;  
RA Iwai Y., Usui T., Hirano S., Steward R., Takeichi M., Uemura T.;  
RT "Axon patterning requires DN-cadherin, a novel neuronal adhesion  
RT receptor, in the Drosophila embryonic CNS."  
RL Neuron 19:77-89(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Broksstein P., Brotlier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferliera S., Fleischmann W.,  
RA Folsler C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,



RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shu B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Glibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [3]  
RP INTERACTION WITH ARM.  
RX MEDLINE-98298928; PubMed-9635189;  
RA Loureiro J., Pelfer M.;  
RT "Roles of Armadillo, a Drosophila catenin, during central nervous  
RT system development.";  
RT Curr. Biol. 8:622-632(1998).  
CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
CC SORTING OF HETEROGENEOUS CELL TYPES. MAY ASSOCIATE WITH ARM NEURAL  
CC ISOFORM AND PARTICIPATE IN THE TRANSMISSION OF DEVELOPMENTAL  
CC INFORMATION.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, THE PROTEIN FIRST APPEARS IN  
CC THE MESODERM AT STAGE 9 AND IS PRESENT IN THE MYOBLASTS AND MUSCLE  
CC FIBERS BY STAGE 12 AND STAGE 14, RESPECTIVELY. AT STAGE 12 THE  
CC PROTEIN IS ALSO LOCATED IN THE AXONS OF THE ENTIRE CNS, BUT NOT IN  
CC THE GLIAL CELLS. IN THIRD INSTAR LARVAE PROTEIN IS EXPRESSED IN  
CC THE CNS NEUROPILE, PHOTORECEPTOR AXONS AND PRECURSORS OF ADULT  
CC MUSCLES.  
CC -1- SIMILARITY: CONTAINS 16 CADHERIN DOMAINS.  
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB002397; BAA22151.1; -.  
DR EMBL; AE003656; AAF53635.1; -.  
DR HSSP; P00740; 11XA.  
DR FlyBase; FBgn0015609; Cadn.  
DR InterPro; IPR002126; Cadherin.  
DR InterPro; IPR000233; Cadherin\_C\_term.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR000742; EGF 2.  
DR InterPro; IPR001881; EGF-Ca.  
DR InterPro; IPR001791; Laminin\_G.  
DR Pfam; PF00028; cadherin; 14.  
DR Pfam; PF01049; Cadherin\_C\_term; 1.  
DR Pfam; PF00008; EGF; 3.  
DR Pfam; PF00054; laminin\_G; 2.  
DR PRINTS; PR00205; CADHERIN.  
DR SMART; SM00112; CA; 16.  
DR SMART; SM00179; EGF\_CA; 1.  
DR SMART; SM00001; EGF\_like; 3.  
DR SMART; SM00282; Lamg; 2.  
DR PROSITE; PS00232; CADHERIN\_1; 9.  
DR PROSITE; PS0268; CADHERIN\_2; 16.  
DR PROSITE; PS00022; EGF\_1; 3.

[illegible]

```

AC Q58897; 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1502.
GN MJ1502.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1 SIMILARITY: BELONGS TO THE ATSA / ELAC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67591; AAB99515.1; -.
DR TIGR; MJ1502; -.
DR InterPro; IPR001279; Beta_Lactam_met.
DR Pfam; PF00753; lactamase_B_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 324 AA; 36893 MW; A3DF62E3BAEEFFA4 CRC64;

Query Match 11.0%; Score 78.5; DB 1; Length 324;
Best Local Similarity 24.2%; Pred. No. 3.6;
Matches 39; Conservative 18; Mismatches 57; Indels 47; Gaps 7;

OY 7 GYKKGDASYFEPTGPLYMNVTVGVDKGNEL-LSPHYVEPIKPGTTLTKETIYYVE 65
   1 : : : 1 1 1 : : 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 87 GFGREKELKIFGPECTKEIIE-----NSLKLGTTHYIEPPIKVEIYTKEPITTYKE 138
   1 : : : 1 1 1 : : 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 66 WALDATAY-----KEFRVELD-PSAKIEVTYYDKNK----- 97
   1 : : : 1 1 1 : : 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 139 ENVEIIAYPTEHGIPSYAVYIFKEIKKPRLDIEKAKKLGKIGPDLKLLKNGEAVKNIYGE 198
   1 : : : 1 1 1 : : 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 98 --KEE-----TKSPITEKGFVVP--DLSEHIKNPGFNLI 128
   1 : : : 1 1 1 : : 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 199 IIKPEYVLLPPKKGFCFLAYSGDTLPLEDFGKYLKELGCDVL 239

RESULT 5
IF39_TOBAC STANDARD; PRT; 719 AA.
AC P56821;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Eukaryotic translation Initiation factor 3 subunit 9 (eIF-3 eta) (eIF3
DE p110).
GN PRT1.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```

```

OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. BRIGHT YELLOW 2;
RA Shen W.H., Gigot C.;
RT "Characterization of Prt1, a gene encoding for one of the subunits of
RT the translation initiation factor 3 (eIF3), from Nicotiana tabacum.";
RL Plant Sci. 143:45-54(1999).
CC -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA.
CC -1- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenselsb-sib.ch).
-----
CC EMBL: Y11996; CAA72721.1; -.
DR DR InterPro; IPR000504; RRM.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
FW Initiation factor; Protein biosynthesis; RNA-binding.
FT DOMAIN 60 147 RNA-BINDING (RRM).
SQ SEQUENCE 719 AA; 82524 MW; BBFA223B7DA52BB2 CRC64;

Query Match 11.0%; Score 78.5; DB 1; Length 719;
Best Local Similarity 26.2%; Pred. No. 9.4;
Matches 33; Conservative 23; Mismatches 41; Indels 29; Gaps 7;

QY 4 FDGKRYKKGDASYF-----EPTGYLMVNVTVGVDKGN-----ELLSPH 43
   || : : | : | : || | : || | : || |
Db 543 FDVDELETMASAEHFMAVDVEMDPTGRVATSVTSVHEMENGFNIMWSFGKLLYRILKDH 602
QY 44 YVEF---PIKPGTTLTKKIEIYVVEWALDATAY-KEFRVVELDPSAKIEVTYYDKNKK-K 98
   : : : | : | : | | | | : | : | : | : | : | : | : | : | : |
Db 603 FFQYLMRP-RPSPFLSKEKEE---EIAKNLKRYSKKYEAEQDVSLQLSEODREKRKKLK 658
QY 99 EETKSF 104
   || : :
Db 659 EEWEMW 664

RESULT 6
TOP5_BPT4
ID TOP5_BPT4 STANDARD; PRT; 442 AA.
AC P07065;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA topoisomerase medium subunit (EC 5.99.1.3) (Protein Gp52).
GN 52.
OS Bacteriophage T4.
OC viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID=10665;
OX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87016377; PubMed=3020513;
RA Huang W.M.;
RT "The 52-protein subunit of T4 DNA topoisomerase is homologous to the
RT gyrase-protein of gyrase.";
RL Nucleic Acids Res. 14:7379-7390(1986).
RN [2]
RP REVISIONS.

```

RA Huang W.M.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,  
 RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;  
 RT "Bacteriophage T4 genome analysis."  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 1-46 FROM N.A.  
 RX MEDLINE=88172481; Pubmed=3280805;  
 RA Chapman D., Morad I., Kaufmann G., Gait M.J., Jorissen L., Snyder L.;  
 RT "Nucleotide and deduced amino acid sequence of stp: the bacteriophage  
 RT T4 antitoxin gene."  
 RL J. Mol. Biol. 199:373-377(1988).  
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT  
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. T4 TOPOISOMERASE  
 CC MAKES DOUBLE-STRAND BREAKS.  
 CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining  
 CC of double-stranded DNA.  
 CC -1- SUBUNIT: THIS IS ONE OF THE 3 SUBUNITS OF T4-DNA TOPOISOMERASE.  
 CC -1- SIMILARITY: TO THE PROKARYOTIC GYRASE SUBUNIT A AND TO EUKARYOTIC  
 CC TOPOISOMERASE II.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X04376; CAA27959.1; -  
 DR EMBL; AF158101; AAD42487.1; -  
 DR PIR; B24705; ITBPT4.  
 DR PIR; S01872; S01872.  
 DR HSSP; P09097; LAB4.  
 DR InterPro; IPR002205; DNA\_topoisomIV.  
 DR Pfam; PF00521; DNA\_topoisomIV; 1.  
 DR SMART; SM00434; TOP4c; 1.  
 KW Isomerase; Topoisomerase; DNA-binding.  
 FT ACT\_SITE 117 117  
 FT 5'-ENDS OF THE DNA VIA A PHOSPHO-TYROSYL  
 FT LINKAGE (POTENTIAL).  
 SQ SEQUENCE 442 AA; 50493 MW; 519E60AE6F75AF7 CRC64;  
 Query Match 10.8%; Score 77.5; DB 1; Length 442;  
 Best Local Similarity 25.5%; Pred. No. 6.5;  
 Matches 36; Conservative 21; Mismatches 43; Indels 41; Gaps 8;  
 OY 16 SYFEPGPLYMVN-VTGVDKSKGNELLSPHYVEPIK-----PGTTLTKEKIEYVEVAL 68  
 DB 148 AFYLPPIPTVLLNGVSGIATGYATYLLPHSVSSVKKAVLQALQKKVTKPKVE----- 200  
 OY 69 DATAKERR--VVELDPSAKIEVY-----YDKNKKKEETKSF-PITEKGF 111  
 DB 201 ----FPFERGEVEIDGQYEIRGTYKFTSRQTMHTEIPYKYDRETYVSKILDPLENGKF 256  
 OY 112 VVPD--LSEHIKNPGFNLTIK 130  
 DB 257 ITWDACGEH---GFGFKVK 273

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (FORM PRL-R3).  
 RC STRAIN=C3H; TISSUE=Mammary gland;  
 RX MEDLINE=94085788; Pubmed=8262385;  
 RA Moore R.C., Oka T.;  
 RT "Cloning and sequencing of the cDNA encoding the murine mammary gland  
 RT long-form prolactin receptor."  
 RL Gene 134:263-265(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (FORM PRL-R3).  
 RC STRAIN=SWISS WEBSTER; TISSUE=Liver;  
 RX MEDLINE=93307149; Pubmed=8319571;  
 RA Clarke D.L., Linzer D.I.H.;  
 RT "Changes in prolactin receptor expression during pregnancy in the  
 RT mouse ovary."  
 RL Endocrinology 133:224-232(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (FORM PRL-R3).  
 RA Sasaki M.;  
 RL Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A. (FORM PRL-R3).  
 RC STRAIN-BALB/C; TISSUE=Mammary gland;  
 RA Edey M., Pezet A., Nandi S., Kelly P.A.;  
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A. (FORMS PRL-R2 AND PRL-R1).  
 RC STRAIN=SWISS WEBSTER; TISSUE=Liver;  
 RX MEDLINE=89261824; Pubmed=2725531;  
 RA Davis J.A., Linzer D.I.H.;  
 RT "Expression of multiple forms of the prolactin receptor in mouse  
 RT liver."  
 RL Mol. Endocrinol. 3:674-680(1989).  
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR THE ANTERIOR PITUITARY HORMONE  
 CC PROLACTIN, AS WELL AS PLACENTAL LACTOGEN I AND II.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; PRL-R1, PRL-R2 AND PRL-R3 (SHOWN  
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L13593; AAC37641.1; -  
 DR EMBL; L14811; AAA02686.1; -  
 DR EMBL; D10214; BAA01066.1; -  
 DR EMBL; X73372; CAA51789.1; -  
 DR EMBL; M22959; AAA39977.1; -  
 DR EMBL; M22958; AAA39976.1; -  
 DR PIR; JT0671; JT0671.  
 DR HSSP; P14787; IAN3.  
 DR MGD; MGI:97763; Ptlr.  
 DR InterPro; IPR002996; CRIA.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR003528; Hematopo\_receptor\_L\_F1.  
 DR Pfam; PF00041; fn3; 2.  
 DR SMART; SM00060; FN3; 1.  
 DR PROSITE; PS01352; HEMATOPO\_REC\_L\_F1; 1.  
 KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat;  
 KW Alternative splicing.  
 FT SIGNAL 1 19  
 FT CHAIN 20 608  
 FT DOMAIN 20 229  
 FT TRANSMEM 230 253  
 FT POTENTIAL.  
 FT PROLACTIN RECEPTOR.  
 FT EXTRACELLULAR (BY SIMILARITY).  
 FT BY SIMILARITY.

```
FT DOMAIN 254 608 CYTOPLASMIC (BY SIMILARITY).
FT DOMAIN 20 117 FIBRONECTIN TYPE-III 1.
FT DOMAIN 119 222 FIBRONECTIN TYPE-III 2.
FT DISULFID 31 41 BY SIMILARITY.
FT DISULFID 70 81 BY SIMILARITY.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 281 292 KGSSELSALG -> VHNKEQLENYVY (IN ISOFORM PRL-R2).
FT VARSPLIC 293 608 MISSING (IN ISOFORM PRL-R2).
FT VARSPLIC 281 303 KGSSELSALGCODEPPTSDCE -> LWCSTLQSLVKI
FT VARSPLIC 304 608 PTEFLCDL (IN ISOFORM PRL-R1).
FT CONFLICT 558 558 MISSING (IN ISOFORM PRL-R1).
FT CONFLICT 558 558 L -> F (IN REF. 2).
SQ SEQUENCE 608 AA; 68240 MW; B8CE202B2EFC9FC6 CRC64;

Query Match 10.8%; Score 77; DB 1; Length 608;
Best Local Similarity 23.9%; Pred. No. 11;
Matches 34; Conservative 17; Mismatches 43; Indels 48; Gaps 7;

QY 4 FDKGKKKKGDASY-----FEPTGPY--LMVNVTVGVDKGNELSP-HYVEFP--I 49
DB 274 FDTLLEKKGKSELSALGCGDFPPTSDCEDLLVEFLEVDNEDERLMPSHSKYEPGGV 333
QY 50 KPGTTLTKETKEIEYVEMALDATAKFEFVVELDPSAKIEVYYDK----NKKKEETKSP 105
DB 334 KP-----THLDPDSDSGHGSYDSSHLSLSEKCEEPQAYP 366
QY 106 ITEKGFVVPDLSEHIKNPGFNL 127
DB 367 ---PAFHPIPETEKPENPEANI 385

RESULT 8
VL1_BPv4 STANDARD; PRT; 506 AA.
AC P08341;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major capsid protein L1.
GN L1.
OS Bovine papillomavirus type 4.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87282264; PubMed=3039043;
RA Patel K.R., Smith K.T., Campo M.S.;
RT "The nucleotide sequence and genome organization of bovine
RT papillomavirus type 4."
RL J. Gen. Virol. 68:2117-2128(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X05817; -; NOT_ANNOTATED_CDS.
DR EMBL: D00146; BAA00101.1; -.
DR PIR: B26214; P1WLB4.
DR InterPro: IPR002210; PV_capsid_L1.
DR Pfam: PF00500; late_protein_L1; 1.
DR PRINTS: PR00865; HPVcapsidL1.
DR PRODOM: PD000544; PV_capsid_L1; 1.
KW Coat protein; Late protein.
SQ SEQUENCE 506 AA; 57902 MW; F1C5DFDB54FA681E CRC64;
```

```
Query Match 10.7%; Score 76.5; DB 1; Length 506;
Best Local Similarity 28.7%; Pred. No. 9.5;
Matches 29; Conservative 19; Mismatches 36; Indels 17; Gaps 7;

QY 29 VTGVDS-KG-NELSPHYVEFPKPGTTLTKETKEIEYV---EWALDATAKFEFVVELD 82
DB 331 VTAVDSTRGTNFSISVHTTDPVEVKPOETTYATKFKHYLRHVEEW-DLSLIMQLCIVNLT 388
QY 83 PSAKIEVTTYDKNKKKEETKSPITEKGFVVP--DLSEHIK 121
DB 389 P-----ESIAYLHNMNESIENMWL--GFIQPPNDIEDHYR 422

RESULT 9
LXB1_PHOLU STANDARD; PRT; 327 AA.
ID LXB1_PHOLU
AC P19840;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Alkanal monooxygenase beta chain (EC 1.14.14.3) (Bacterial luciferase
DE beta chain).
GN LUXB.
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photobacterium.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29999;
RX MEDLINE=90375532; PubMed=2204626;
RA Sztitner R., Meighen E.;
RT "Nucleotide sequence, expression, and properties of luciferase coded
RT by lux genes from a terrestrial bacterium."
RL J. Biol. Chem. 265:16581-16587(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29999;
RX MEDLINE=90343746; PubMed=2383248;
RA Johnston T.C., Rucker E.B., Cochrum L., Hruska K.S., Vandegrift V.;
RT "The nucleotide sequence of the luxA and luxB genes of Xenorhabdus
RT luminescens HM and a comparison of the amino acid sequences of
RT luciferases from four species of bioluminescent bacteria."
RL Biochem. Biophys. Res. Commun. 170:407-415(1990).
CC -1- FUNCTION: LIGHT-EMITTING REACTION IN LUMINOUS BACTERIA. THE
CC SPECIFIC ROLE OF THE BETA SUBUNIT IS UNKNOWN, BUT IT IS ABSOLUTELY
CC REQUIRED FOR BIOLUMINESCENCE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: RCHO + FMN(2) + O(2) = RCOOH + FMN + H(2)O +
CC light.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M57416; AAA27624.1; -.
DR EMBL: M90093; AAA27620.1; -.
DR EMBL: M55977; AAA27627.1; -.
DR PIR: C37898; C37898.
```



DR HSSP; P07739; 1BSL.  
DR InterPro; IPR002103; Bac\_luciferase.  
DR Pfam; PF00296; bac\_luciferase; 1.  
DR PRINTS; PR00089; LUCIFERASE.  
DR PROSITE; PS00494; BACTERIAL\_LUCIFERASE; 1.  
KW Photoprotein; Luminescence; Oxidoreductase; Monooxygenase;  
FT Flavoprotein; FMN.  
FT VARIANT 85 85 A -> R (IN STRAIN HM).  
SQ SEQUENCE 327 AA; 37595 MW; 1905AF01658BE56E CRC64;

Query Match 10.6%; Score 75.5; DB 1; Length 327;  
Best Local Similarity 25.4%; Pred. No. 7;

Matches 33; Conservative 18; Mismatches 44; Indels 35; Gaps 7;

QY 18 FEPTGPYLMVNTGVDKGNELSPHYVEFPKPGTTL-----TKEIEYEWALDA 70  
Db 163 YFPGGPRKYVTAT-----SHIIVEMAAKKGIFLFWKDDSDNDRYEYAEYKAVA 212  
QY 71 TAYKEFRVVELDPSAKIEVY-YDKNKKKEETKSPITEKGFVVPDLSEHI-----KNPGF 125  
Db 213 DKY-DVDLSEIDHQLMILVYNEDSNKAKQETRAF-----ISDYVLEHMPNENF 260  
QY 126 -NLITKVIE 134  
Db 261 ENKLEETIAE 270

RESULT 10

NAE2\_THEME STANDARD; PRT; 576 AA.  
ID NAE2\_THEME

AC Q9X0Y0;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable glutamine-dependent NAD(+) synthetase (EC 6.3.5.1) (NAD(+)  
DE synthase [glutamine-hydrolysing]).  
GN NADE2 OR TM1253.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogales; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Utterback T.R., Pratt M.S., Phillips C.A., Garrett M.M.,  
RA Stewart A.M., Sutton M.D., Pratt M.S., Fleischmann R.D., Eisen J.A., White O.,  
RA Heidelberg J., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
RT genome sequence of Thermotoga maritima.";  
RL Nature 399:323-329(1999).

CC -1- FUNCTION: CAN USE BOTH GLUTAMINE OR AMMONIA AS A NITROGEN  
CC SOURCE (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP + deamido-NAD(+) + L-glutamine + H(2)O =  
CC AMP + diphosphate + NAD(+) + L-glutamate.  
CC -1- PATHWAY: DE NOVO BIOSYNTHESIS OF NAD.  
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE NAD  
CC SYNTHETASE FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AE001780; AAD36328.1; -.  
DR TIGR; TM1253; -.  
DR InterPro; IPR003010; CN\_hydrolase.

DR InterPro; IPR003694; NAD\_synthase.  
DR InterPro; IPR000132; Nitric\_cyn\_hydratase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PRINTS; PR002540; NAD\_synthase; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; UNKNOWN\_1.  
KW ligase; NAD; ATP-binding; Complete proteome.  
FT DOMAIN 292 576 LIGASE.  
FT NP\_BIND 321 328 ATP (BY SIMILARITY).  
FT ACT\_SITE 323 323 BY SIMILARITY.  
SQ SEQUENCE 576 AA; 64919 MW; 208FDC77964C957F CRC64;

Query Match 10.5%; Score 75; DB 1; Length 576;  
Best Local Similarity 18.1%; Pred. No. 15;

Matches 39; Conservative 31; Mismatches 54; Indels 92; Gaps 7;

QY 7 GKYYK-----GDDASYEPTGPYLMVNT-----TGV 32  
Db 108 GYRKISLPNYGVDERRYFKPGEEELLVKIGNIKGVITICEDIMNPVPSASLSGEGV 167  
QY 33 DSKGNELSPHYVEFPKPGTTLTKEIEYEWMA-----LDATA- 72  
Db 168 HLIANLSASPHYGKPVLRKDYLSMKAYDYHAYACNMVGGODELVFDGSGMVDASGE 227  
QY 73 -----YKEFRVVELDPSAKIEVYDYDKNKKKEETKSPIT--TEKG----- 110  
Db 228 VINYKGLFEEELITVDLIDENLRVSLVDPRRRYMKQTQNPVKYVEAGNLREKSGHEPV 287  
QY 111 -----FVVPDLSEHIKNPGFNLTQVI 133  
Db 288 VNPLPVREEMERFALITGLRDYVRKNGFE---KVVI 320

RESULT 11

FAS\_CHICK STANDARD; PRT; 2511 AA.  
ID FAS\_CHICK

AC P12276;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Fatty acid synthase (EC 2.3.1.85) [includes: EC 2.3.1.38; EC 2.3.1.39;  
DE EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].  
GN FASN OR FAS.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.

OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.  
RC STRAIN=WHITE LEGHORN; TISSUE=Liver;  
RX MEDLINE=95031085; PubMed=7944406;  
RA Huang W.-Y., Chitrala S.S., Wakil S.J.;  
RT "Amino-terminal blocking group and sequence of the animal fatty acid  
RT synthase.";  
RL Arch. Biochem. Biophys. 314:45-49(1994).

RN [2]  
RP SEQUENCE OF 75-1775 FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=89282777; PubMed=2734291;  
RA Holzer K.P., Liu W., Hammes G.G.;  
RT "Molecular cloning and sequencing of chicken liver fatty acid  
RT synthase cDNA.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:4387-4391(1989).

RN [3]  
RP SEQUENCE OF 1568-2512 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=89139426; PubMed=2917973;  
RA Chitrala S.S., Kasturi R., Pazirandeh M., Stolow D.T., Huang W.-Y.,  
RA Wakil S.J.;  
RT "A novel cDNA extension procedure. Isolation of chicken fatty acid  
RT synthase cDNA clones.";  
RL J. Biol. Chem. 264:3750-3757(1989).  
RN [4]



RP SEQUENCE OF 1752-2512 FROM N.A.  
RX MEDLINE-88320436; PubMed=2842766;  
RA Yuan Z., Liu W., Hammes G.G.;  
RT "Molecular cloning and sequencing of DNA complementary to chicken  
RT liver fatty acid synthase mRNA.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:6328-6331(1988).  
RN [5]  
RP SEQUENCE OF 2202-2512 FROM N.A.  
RX MEDLINE-89088152; PubMed=3207710;  
RA Kasturi R., Chirala S.S., Pazirandeh M., Wakil S.J.;  
RT "Characterization of a genomic and cDNA clone coding for the  
RT thioesterase domain and 3' noncoding region of the chicken liver  
RT fatty acid synthase gene.";  
RL Biochemistry 27:7778-7785(1988).  
RN [6]  
RP SEQUENCE OF 2121-2209.  
RX MEDLINE-89192401; PubMed=2648999;  
RA Huang W.-Y., Stoops J.K., Wakil S.J.;  
RT "Complete amino acid sequence of chicken liver acyl carrier protein  
RT derived from the fatty acid synthase.";  
RL Arch. Biochem. Biophys. 270:92-98(1989).  
RN [7]  
RP SEQUENCE OF 2209-2508.  
RC STRAIN-WHITE LECHORN;  
RX MEDLINE-89088151; PubMed=3207709;  
RA Yang C.-Y., Huang W.-Y., Chirala S.S., Wakil S.J.;  
RT "Complete amino acid sequence of the thioesterase domain of chicken  
RT liver fatty acid synthase.";  
RL Biochemistry 27:7773-7777(1988).  
RN [8]  
RP SEQUENCE OF 667-674 AND 1698-1709.  
RX MEDLINE-89323081; PubMed=2751995;  
RA Chang S.I., Hammes G.G.;  
RT "Amino acid sequences of pyridoxal 5'-phosphate binding sites and  
RT fluorescence resonance energy transfer in chicken liver fatty acid  
RT synthase.";  
RL Biochemistry 28:3781-3788(1989).  
CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF  
CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.  
CC THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN  
CC ACYL CARRIER PROTEIN.  
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a  
CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).  
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +  
CC acetyl-[acyl-carrier protein].  
CC -1- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +  
CC malonyl-[acyl-carrier protein].  
CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-  
CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +  
CC [acyl-carrier protein].  
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +  
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.  
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]  
CC = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.  
CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-  
CC 2,3-dehydroacyl-[acyl-carrier protein] + NADPH.  
CC -1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-  
CC carrier protein] + oleate.  
CC -1- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.  
CC -1- CAUTION: REF.3 AND REF.5 SEQUENCES DIFFER FROM THAT SHOWN FROM  
CC POSITION 2351 ONWARDS DUE TO A FRAMESHIFT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J04485; AAB46389.1; -  
DR EMBL: J03860; AAA48767.1; -  
DR EMBL: J02839; AAA82106.1; ALT\_SEQ.

DR PIR: A33918; XYCHEA.  
DR PIR: A32015; A32015.  
DR HSSP: P02901; 1ACP.  
DR InterPro: IPR001227; Acyltransferase\_domain.  
DR InterPro: IPR002085; Adh\_zn\_family.  
DR InterPro: IPR000794; Ketoacyl-synt.  
DR InterPro: IPR003880; Phosphopant\_attach.  
DR InterPro: IPR001031; Thioesterase.  
DR Pfam: PF00698; Acyl\_transfer\_1.  
DR Pfam: PF00107; adh\_zinc\_1.  
DR Pfam: PF00109; ketoacyl-synt\_1.  
DR Pfam: PF02801; ketoacyl-synt\_C\_1.  
DR Pfam: PF00550; pp-binding\_1.  
DR Pfam: PF00975; Thioesterase\_1.  
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.  
DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE; 1.  
DR PROSITE: PS50075; ACP\_DOMAIN; 1.  
DR Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;  
DR Transferase; Hydrolase; Oxidoreductase; Lyase; NADP; Acetylation;  
DR Alternative splicing; Pyridoxal phosphate.  
DR INIT\_MET 0  
DR DOMAIN 0  
DR DOMAIN 1  
DR DOMAIN 427  
DR DOMAIN 1638  
DR DOMAIN 1867  
DR DOMAIN 2124  
DR DOMAIN 2209  
DR MOD\_RES 1  
DR ACT\_SITE 160  
DR ACT\_SITE 579  
DR ACT\_SITE 579  
DR ACT\_SITE 877  
DR ACT\_SITE 877  
DR NP\_BIND 1674  
DR BINDING 1707  
DR NP\_BIND 1888  
DR BINDING 2157  
DR ACT\_SITE 2308  
DR ACT\_SITE 2481  
DR VARSPLIC 2348  
DR CONFLICT 77  
DR CONFLICT 116  
DR CONFLICT 675  
DR CONFLICT 1169  
DR CONFLICT 1178  
DR CONFLICT 1191  
DR CONFLICT 1198  
DR CONFLICT 1286  
DR CONFLICT 1372  
DR CONFLICT 1533  
DR CONFLICT 1577  
DR CONFLICT 1685  
DR CONFLICT 1732  
DR CONFLICT 1745  
DR SEQUENCE 2511 AA; 274648 MW; 622039DAC8315D3F CRC64;

Query Match 10.4%; Score 74.5; DB 1; Length 2511;  
Best Local Similarity 25.5%; Pred. No. 98;  
Matches 26; Conservative 13; Mismatches 24; Indels 39; Gaps 5;  
QY 27 VNVTVGDSKGNELLSPHYVERPIKPGTTLTKKEIEYVEMALDATAKYKFRVVELDPSAK 86  
DB 801 IHLTGIVLGNLFP--VEYVPVGTPL-----ISPIYKW-----DHSQD 839  
QY 87 IEVTVYDKNKKKEETKSPITEKGF-----VVPDLSEH 119  
DB 840 WDV-----PKAEDFPGSGKSASASVYNIIDVSPDSPDH 872

RESULT 12  
MOAA\_METJA

```
ID MOAA_METUA STANDARD: PRT; 298 AA.
AC 058234;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative molybdopterin cofactor synthesis protein A.
GN MOAA OR M0824.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.;
RL Science 273:1058-1073(1996).
CC -1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF MOLYBDOPTERIN PRECURSOR
CC 2 FROM GUANOSINE (BY SIMILARITY).
CC -1- PATHWAY: MOLYBDENUM COFACTOR BIOSYNTHESIS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MOAA / NIFB / PQOE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67526; AAB98823.1; -.
DR TIGR: M0824; -.
DR InterPro: IPR000385; MoAA_NifB_Pqoe.
DR Pfam: PF01444; MoAA_NifB_Pqoe; 1.
DR PROSITE: PS01305; MoAA_NifB_PQOE; 1.
KW Molybdenum cofactor biosynthesis; Iron-sulfur; Complete proteome.
FT METAL 20 20 IRON-SULFUR (POTENTIAL).
FT METAL 24 24 IRON-SULFUR (POTENTIAL).
FT METAL 27 27 IRON-SULFUR (POTENTIAL).
SQ SEQUENCE 298 AA; 34652 MW; EA92B9A0A75BFD55 CRC64;

Query Match 10.3%; Score 74; DB 1; Length 298;
Best Local Similarity 21.4%; Pred. No. 8.6;
Matches 21; Conservative 19; Mismatches 42; Indels 16; Gaps 1;

QY 31 GYDSKGNELSPHYVEFPPIKPGTTLTKKIEYYEWALDATAVKEFRVVELDPSAKIEVT 90
| | | : : : | | : : : |
DB 31 GHDSNDRYMTPEIGIIAKISTKFGVKI-----KISGEPILRKDVC 74

QY 91 YVDKNNKKEETKSPITEKGFVVPDLSEHIKPGFNLI 128
| | | | | : : : | : : : | : : : |
DB 75 EIIENIKDERIKDISLTNGIILENLAEKLDAGLNRV 112

RESULT 13
V234_FOWPV STANDARD: PRT; 428 AA.
ID V234_FOWPV
AC P14368; P14367; Q9J501;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative ankylrin-repeat protein FPV234 (BamHI-ORF12/ORF13).
GN FPV234.
```

```
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20193820; PubMed=10729156;
RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus."
RL J. Virol. 74:3815-3831(2000).
RN [2]
RP SEQUENCE OF 65-428 FROM N.A.
RC STRAIN=FP-9 / Isolate HP-438;
RX MEDLINE=88229622; PubMed=2836548;
RA Tomley F., Bins M., Campbell J., Boursnell M.E.G.;
RT "Sequence analysis of an 11.2 kilobase, near-terminal, BamHI fragment
RT of fowlpox virus."
RL J. Gen. Virol. 69:1025-1040(1988).
CC -1- SIMILARITY: CONTAINS 8 ANK REPEATS.
CC -1- CAUTION: Ref 2 sequence differs from that shown due to frameshifts
CC in position 204, 219 and 237.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF198100; AAF44578.1; -.
DR EMBL: D00295; BAA00209.1; ALT_FRAME.
DR EMBL: D00295; BAA00207.1; ALT_FRAME.
DR PIR: C30087; WMVZTW.
DR PIR: D30087; WMVZTH.
DR InterPro: IPR002110; ANK.
DR Pfam: PF00023; ank; 4.
DR SMART: SM00248; ANK; 6.
DR PROSITE: PS50088; ANK_REPEAT; 5.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
KW Hypothetical protein; Repeat; ANK repeat.
FT REPEAT 6 35 ANK 1.
FT REPEAT 39 68 ANK 2.
FT REPEAT 71 100 ANK 3.
FT REPEAT 103 132 ANK 4.
FT REPEAT 137 169 ANK 5.
FT REPEAT 174 202 ANK 6.
FT REPEAT 206 238 ANK 7.
FT REPEAT 242 271 ANK 8.
SQ SEQUENCE 428 AA; 49194 MW; C71BEAF74687EE8F CRC64;

Query Match 10.3%; Score 74; DB 1; Length 428;
Best Local Similarity 26.6%; Pred. No. 13;
Matches 29; Conservative 14; Mismatches 38; Indels 28; Gaps 4;

QY 43 HYVEFPPIKPGTTLTKKI-EYVEMALDATA-----YKEFRV---VEL 81
| | | : : : | : : : | : : : |
DB 102 HYCLKPIHIAANRTESKIVKLLIEYGADINSEDEGANGKYPPIHYAMKYDPRFKIIVLL 161

QY 82 DPSAKIEVTVYDKNKKKEETKSPITEKGFVVPDLSEHIKPGFNLI 130
| | | | | : : : | : : : | : : : |
DB 162 DHGA-----DINKQSVLTNTSPLYETRTITDLDLYIISRGANINIK 203

RESULT 14
PRLR_RAT STANDARD: PRT; 610 AA.
ID PRLR_RAT
AC P05710; Q63451; Q63723; Q62832; Q64274; Q63479;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Prolactin receptor precursor (PRL-R) (Lactogen receptor).
GN
```

GN PRLR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91155946; PubMed=2293022;  
 RA Shirota M., Banville D., All S., Jolicoeur C., Boutin J.M.,  
 RA Ederly M., Djane J., Kelly P.A.;  
 RT "Expression of two forms of prolactin receptor in rat ovary and  
 RT liver.";  
 RL Mol. Endocrinol. 4:1136-1143(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A. (LONG FORM AND SHORT FORM).  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=OVARY;  
 RX MEDLINE=90241201; PubMed=2159291;  
 RA Zhang R., Buczek E., Tsai-Morris C.H., Hu Z.Z., Dufau M.L.;  
 RT "Isolation and characterization of two novel rat ovarian lactogen  
 RT receptor cDNA species.";  
 RL Biochem. Biophys. Res. Commun. 168:415-422(1990).  
 RN [3]  
 RP SEQUENCE OF 281-610 FROM N.A.  
 RA Banville D., Stocco R., Murthy K.K., Bole Y., Kelly P.A.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (MEDIUM FORM).  
 RC TISSUE=Liver;  
 RX MEDLINE=88165059; PubMed=2832068;  
 RA Boutin J.-M., Jolicoeur C., Okamura H., Gagnon J., Ederly M.,  
 RA Shirota M., Banville D., Dusanter-Fourt I., Djane J., Kelly P.A.;  
 RT "Cloning and expression of the rat prolactin receptor, a member of  
 RT the growth hormone/prolactin receptor gene family.";  
 RL Cell 53:69-77(1988).  
 RN [5]  
 RP SEQUENCE FROM N.A. (FORM NB2).  
 RC TISSUE=Lymphoma;  
 RX MEDLINE=92041834; PubMed=1718958;  
 RA All S., Pelligrini I., Kelly P.A.;  
 RT "A prolactin-dependent immune cell line (Nb2) expresses a mutant form  
 RT of prolactin receptor.";  
 RL J. Biol. Chem. 266:20110-20117(1991).  
 RN [6]  
 RP SEQUENCE FROM N.A. (FORM NB2).  
 RX MEDLINE=95014432; PubMed=7929319;  
 RA O'Neal K.D., Yu-Lee L.Y.;  
 RT "Differential signal transduction of the short, Nb2, and long  
 RT prolactin receptors. Activation of interferon regulatory factor-1 and  
 RT cell proliferation.";  
 RL J. Biol. Chem. 269:26076-26082(1994).  
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR THE ANTERIOR PITUITARY HORMONE  
 CC PROLACTIN.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS ARE PRODUCED BY ALTERNATIVE  
 CC SPLICING OF THE PRLR GENE.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -----  
 CC This swiss-prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M57668; AAA41938.1; -  
 DR EMBL; M34083; AAA79273.1; -  
 DR EMBL; L48060; AAA79274.1; -  
 DR EMBL; U34730; AAA92053.1; -  
 DR EMBL; M19304; AAA41937.1; -  
 DR EMBL; M74152; AAA41946.1; -  
 DR EMBL; U07567; AAA61784.1; -  
 RN

DR PIR; A29884; A29884.  
 DR HSSP; P14787; IAN3.  
 DR InterPro; IPR002996; CRIA.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR003528; Hematopo\_receptor\_L\_F1.  
 DR Pfam; PF00041; fn3; 2.  
 DR SMART; SM00060; FN3; 1.  
 DR PROSITE; PS01352; HEMATOPO\_REC\_L\_F1; 1.  
 KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat;  
 KW Alternative splicing.  
 FT SIGNAL 1 19  
 FT CHAIN 20 610  
 FT DOMAIN 20 229  
 FT TRANSMEM 230 253  
 FT DOMAIN 254 610  
 FT DOMAIN 20 117  
 FT DOMAIN 119 222  
 FT DISULFID 31 41  
 FT DISULFID 70 81  
 FT DISULFID 54 54  
 FT CARBOHYD 99 99  
 FT CARBOHYD 127 127  
 FT VARSPLIC 131 150  
 FT VARSPLIC 151 610  
 FT VARSPLIC 281 310  
 FT VARSPLIC 311 610  
 FT VARSPLIC 342 539  
 FT CONFLICT 236 236  
 FT CONFLICT 345 345  
 FT CONFLICT 465 465  
 FT CONFLICT 466 466  
 FT CONFLICT 469 469  
 FT CONFLICT 541 541  
 FT CONFLICT 555 555  
 SQ SEQUENCE 610 AA; 68599 MW; 83D04D832861295D CRC64;  
 Query Match 10.3%; Score 74; DB 1; Length 610;  
 Best Local Similarity 23.9%; Pred. No. 20;  
 Matches 34; Conservative 17; Mismatches 43; Indels 48; Gaps 7;  
 QY 4 FDKGKTKKGDASY-----FETGPR--LMVNTGVDSKGNELSP-HYVEFP--I 49  
 DB 274 FDTLLEKKGSEELLSALGQDFPPTSDCEDLLVEFLVDNEDERLMPSHSKYPGQGV 333  
 QY 50 KPPTTLTKKIEIYVWALDATAYKKEPRVELDPSAKIEVYYDK---NKKKEETKSP 105  
 DB 334 KP-----THLDPDSDSGHSYSHLSLSEKCEEPQAYP 366  
 QY 106 ITEKGFVVPDISEHIKNPGFNL 127  
 DB 367 PT--LHPEITEKPEPEANI 385  
 RESULT 15  
 SERC\_ARATH  
 ID SERC\_ARATH STANDARD; PRT; 430 AA.  
 AC Q96255;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Phosphoserine aminotransferase, chloroplast precursor (EC 2.6.1.52)  
 DE (PSAT).  
 GN AT4G35630 OR F8D20.140.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]



RP SEQUENCE FROM N.A.  
RX MEDLINE=99097839; PubMed=9881164;  
RA Ho C.-L., Noji M., Saito M., Yamazaki M., Saito K.;  
RT "Molecular characterization of plastidic phosphoserine  
amino transferase in serine biosynthesis from Arabidopsis.";  
RL Plant J. 16:443-452(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083488; PubMed=10617198;  
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,  
RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,  
RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,  
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,  
RA Reichert B., Portetelie D., Perez-Alonso M., Boutry M., Bancroft I.,  
RA vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,  
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
RA van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,  
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
RA Weltzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
RA Bernelsier S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,  
RA van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,  
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,  
RA Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,  
RA Dose S., de Haan M., Maarsee A.C., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,  
RA Neumann S., Argirion A., Vitale D., Liguori R., Piravandi E.,  
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,  
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
RA Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bieleke C.,  
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
RA Nelson J., Spieth J., Ryan E., Andrews S., Gelsel C., Layman D.,  
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
RA Ma P., Zhong J., Preston R., Vili D., Shekher M., Matero A., Shah R.,  
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
RA Chen E., Marra M., Martienssen R., McCombie W.R.;  
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
thaliana.";  
RL Nature 402:769-777(1999).  
CC -1- CATALYTIC ACTIVITY: O-phospho-L-serine + 2-oxoglutarate = 3-  
CC phosphonooxypyruvate + L-glutamate.  
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -1- PATHWAY: MAJOR PHOSPHORYLATED PATHWAY OF SERINE BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: Chloroplast.  
CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D88541; BAA13640.1; -

```
DR EMBL; AB010408; BAA24441.1; -.
DR EMBL; AL031135; CAA20033.1; -.
DR EMBL; AL161587; CAB80279.1; -.
DR HSSP; p23721; 1BJO.
DR InterPro; IPR000192; AminoTransf_class_V.
DR InterPro; IPR003248; Phosphser_amintransf.
DR Pfam; PF00266; aminotran_5; 1.
DR ProDom; PD001544; Phosphser_amintransf; 1.
DR PROSITE; PS00595; AA_TRANSFER_CLASS_5; 1.
KW Serine biosynthesis; transferase; AminoTransferase;
KW Pyridoxal phosphate; Chloroplast; Transit peptide.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT CHAIN ? 430 PHOSPHOSERINE AMINOTRANSFERASE.
FT BINDING 265 265 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 430 AA; 47359 MW; B3225CC8DE78BDE2 CRC64;.

Query Match 10.3%; Score 73.5; DB 1; Length 430;
Best Local Similarity 22.1%; Pred. No. 15;
Matches 33; Conservative 17; Mismatches 48; Indels 51; Gaps 4;

QY 11 KGDDASYEPTGPLYLMVNTGVDSKGNELLSPHYVEFPKPGTLLTKETIEY---YVEWA 67
   ||| | || :
Db 158 KSDDTVDFVTGTSM-----GDRAVKEAKKYCKTNVIWS 190

QY 68 LDATAYKE--FRVELDPSAKIEVTTYDKNKKKETKSFPITEKGVPDLSEHI---- 120
   :| :| :||| ::|| :| :| :||| :| :| :
Db 191 GKSEKYTKVPSPFEELQTPDAKYLIHICANETIHGEVFEDYPYPKNGFLVADMSNFCSKP 250

QY 121 -----KNPGFNLTQKVIEK 135
      ||| :||| :|
Db 251 VDYSKEGVITYGAQKNVGPSGVTVIIRK 279

RESULT 16
CTRA_CAUCR
ID CTRA_CAUCR STANDARD; PRT; 231 AA.
AC Q45994;
DC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cell cycle transcriptional regulator ctra (Response regulator soka).
GN CTRA OR SOKA OR CC3035.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
ON [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CB15N / NA1000;
RX MEDLINE=96140642; PubMed=8548829;
RA Quon K.C., Marczyński G.T., Shapiro L.;
RT "Cell cycle control by an essential bacterial two-component signal
   transduction protein.";
RL Cell 84:83-93(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Niernan W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Pladke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Winn M.L., Haft D.H.,
RA Kolonay J.F., Smitt J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RN [3]
RP SEQUENCE OF 159-231 FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=98132609; PubMed=9465034;
RA Wu J., Ohta N., Newton A.;
```

RT "An essential, multicomponent signal transduction pathway required for  
RT cell cycle regulation in *Caulobacter*.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:1443-1448(1998).  
CC -1- FUNCTION: FORMS PART OF A TWO-COMPONENT REGULATORY SYSTEM  
CC CTRA/CCKA THAT CONTROLS MULTIPLE EVENTS IN THE CELL CYCLE,  
CC INCLUDING CELL DIVISION, STALK SYNTHESIS, AND CELL CYCLE-SPECIFIC  
CC TRANSCRIPTION. BINDS TO A GROUP OF CELL CYCLE-REGULATED PROMOTERS  
CC CRITICAL FOR DNA REPLICATION, DNA METHYLATION, AND CLASS II  
CC FLAGELLAR BIOGENESIS.  
CC -1- PTM: PHOSPHORYLATED BY CCKA.  
CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U39559; AAA93080.1; .  
DR EMBL; AE005966; AAK24997.1; .  
DR EMBL; AF021339; AAC05479.1; .  
DR HSSP; P08402; 1B00.  
DR TIGR; CC3035; .  
DR InterPro; IPR001789; response\_reg.  
DR InterPro; IPR001867; Trans\_reg\_C.  
DR Pfam; PF00072; response\_reg; 1.  
DR Pfam; PF00486; trans\_reg\_C; 1.  
DR SMART; SM00448; REC; 1.  
DR PROSITE; PS00110; RESPONSE\_REGULATORY; 1.  
DR Sensory transduction; Phosphorylation; Transcription regulation;  
KW DNA-binding; Complete proteome.  
FT DOMAIN 1 116 RESPONSE REGULATORY.  
FT MOD\_RES 51 51 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 231 AA; 25796 MW; BB812AF6EDADB8 CRC64;  
  
Query Match 10.2%; Score 73; DB 1; Length 231;  
Best Local Similarity 26.2%; Pred. No. 7.9;  
Matches 32; Conservative 19; Mismatches 39; Indels 32; Gaps 5;  
  
QY 10 KKGD----DASYEPTGPLYMNVTVGVDKGNELLSPHYVEFPKPGTTLTKKEIEYV 64  
Db 128 KTGDIVNLDKATVEYNGN--RVHLTGKGYMLELLS-----LRKGTTLTKEMFLNHL 178  
QY 65 EWALDATAYKEFRVVELDPSAKIEVYYDKNKKKEETKSF-----PITEKGFVVPDLSE 118  
Db 179 YCGMD-----EPELKIIVDFICKLRKKLASAHGKHHTETWGRGYVLRDPNE 226  
QY 119 HI 120  
Db 227 QY 228  
  
RESULT 17  
TRMD\_BUCAI STANDARD; PRT; 237 AA.  
AC P57476;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE tRNA (Guanine-N1)-methyltransferase (EC 2.1.1.31) (MIG-  
DE methyltransferase) (tRNA [GM37] methyltransferase).  
GN TRMD OR BU396.  
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
OS symbiotic bacterium).  
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
OX NCBI\_TaxID=118099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TOKYO 1998;  
RX MEDLINE-20445173; PubMed-10993077;

RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
RT "Genome sequence of the endocellular bacterial symbiont of aphids  
RT Buchnera sp. APS.";  
RL Nature 407:81-86(2000).  
CC -1- FUNCTION: SPECIFICALLY METHYLATES GUANOSINE-37 IN VARIOUS TRNAS  
CC (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA - S-adenosyl-L-  
CC homocysteine + tRNA containing N1-methylguanine.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -1- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRMD FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AP001119; BAB13099.1; .  
DR InterPro; IPR002649; tRNA\_m1G\_MT.  
DR Pfam; PF01746; tRNA\_m1G\_MT; 1.  
DR ProDom; PD004978; tRNA\_m1G\_MT; 1.  
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.  
SQ SEQUENCE 237 AA; 27335 MW; 86CE3836B35EC4AB CRC64;

Query Match 10.2%; Score 73; DB 1; Length 237;  
Best Local Similarity 25.5%; Pred. No. 8.1;  
Matches 36; Conservative 17; Mismatches 40; Indels 48; Gaps 8;  
  
QY 3 SFDKGYKKGGDASY-----FEPTGPLYM-----NVTGV--DSKGNELLSPH 43  
Db 49 NFSKNKYKSVDDRPYGGGPGMLMSFEPL--YLAIQAKSTSKNTVIYLSPOGKELKQNH 106  
QY 44 YVEPIKPGTTLTKKEIEYV-----EWALDA---TAYKEFRVVELDP 83  
Db 107 IEELLVK-----KKKIIVICGRYEGIDRIIDNOVDEWISGISYILTGGLAAMWIDA 160  
QY 84 SAKIEVYYDKNKKKEETKSF 104  
Db 161 ISRL-IPGVIKTKKSIEDSF 180  
  
RESULT 18  
YPIB\_ECOLI STANDARD; PRT; 263 AA.  
ID YPIB\_ECOLI  
AC P76612;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein ypiB.  
GN YPIB OR B2649.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474(1997).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial





```
FT ACT_SITE 161 161 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 581 581 MALONYLTRANSFERASE (BY SIMILARITY).
FT NP_BIND 1665 1682 NADP (ER).
FT BINDING 1698 1698 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT NP_BIND 1765 1780 NADP (KR).
FT BINDING 2151 2151 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT ACT_SITE 2302 2302 THIOESTERASE (BY SIMILARITY).
FT ACT_SITE 2475 2475 THIOESTERASE (BY SIMILARITY).
FT ACT_SITE 878 878 BETA-HYDROXYACYL DEHYDRATASE (BY SIMILARITY).
FT CONFLICT 871 871 S -> P (IN REF. 3).
FT CONFLICT 1967 1968 MV -> IL (IN REF. 5).
FT CONFLICT 2085 2085 C -> P (IN REF. 4).
FT CONFLICT 2106 2106 A -> V (IN REF. 1 AND 5).
FT CONFLICT 2296 2296 Y -> H (IN REF. 1 AND 5).
SQ SEQUENCE 2505 AA; 272647 MW; 5810EC13D37F3114 CRC64;

Query Match 10.2%; Score 73; DB 1; Length 2505;
Best Local Similarity 30.0%; Pred. No. 1.3e+02;
Matches 24; Conservative 14; Mismatches 32; Indels 10; Gaps 4;

QY 27 VNVTVGDSKGNELSPHYVEFPKPTLTLEKEIYVEMALDAPKPRVVELDP--S 84
   |||:| | | | | | | | | | | | | | | | | | | | | | | | |
DB 803 VHLTGIDINPNALFPP--VEFPVPRGTPPL---ISPHIKW--DHSQTWDIPVAEDFPNGS 854

QY 85 AKIEVTYDKNKKKEETKSF 104
   | | | | | | | | | | | | | | | | | | | | | | | | |
DB 855 SSSSATVYNIDASSESSSDHY 874

RESULT 20
R2_RABIT STANDARD; PRT; 4969 AA.
ID R2_RABIT STANDARD; PRT; 4969 AA.
AC P30957;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ryanodine receptor 2 (Cardiac muscle-type ryanodine receptor) (RyR2)
DE (RyR-2) (Cardiac muscle ryanodine receptor-calcium release channel).
GN RYR2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart muscle;
RX MEDLINE=90337947; PubMed=2380170;
RA Otsu K., Willard H.F., Khanna V.K., Zorzato F., Green N.M.,
RA MacLennan D.H.;
RT "Molecular cloning of cDNA encoding the Ca2+ release channel
RT (ryanodine receptor) of rabbit cardiac muscle sarcoplasmic
RT reticulum.";
RL J. Biol. Chem. 265:13472-13483(1990).
RN [2]
RP PHOSPHORYLATION OF SER-2809.
RX MEDLINE=91250425; PubMed=1645727;
RA Wichter D.R., Kovacs R.J., Schulman H., Cefali D.C., Jones L.R.;
RT "Unique phosphorylation site on the cardiac ryanodine receptor
RT regulates calcium channel activity.";
RL J. Biol. Chem. 266:11144-11152(1991).
CC -1- FUNCTION: Communication between transverse-tubules and
CC sarcoplasmic reticulum. Contraction of cardiac muscle is triggered
CC by release of calcium ions from SR following depolarization of T-
CC tubules.
CC -1- SUBUNIT: Homotetramer (Potential).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: HEART AND BRAIN.
CC -1- MISCELLANEOUS: The calcium release channel is modulated by calcium
CC ions, magnesium ions, ATP and calmodulin.
CC -1- MISCELLANEOUS: The calcium release channel activity resides in the
CC C-terminal region while the remaining part of the protein
```

```
CC constitutes the 'foot' structure spanning the junctional gap
CC between the SR and the T-tubule. It is possible that the foot
CC structure interacts with the cytoplasmic region of the
CC dihydropyridine receptor.
CC -1- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-
CC release channel in junctional SR and modulates its activity.
CC -1- SIMILARITY: BELONGS TO THE RYANODINE RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M59743; AAA31179.1; -.
DR PIR: A37113; A37113.
DR InterPro: IPR000636; Cation_chan_non_lig.
DR InterPro: IPR001682; Channel_pore_Ca_Na.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR003608; MIR.
DR InterPro: IPR000699; RYDR_ITPR.
DR InterPro: IPR003032; RYR.
DR InterPro: IPR001215; Ryanodn_receptor.
DR InterPro: IPR003877; SPRY.
DR InterPro: IPR003878; SPRY_domain.
DR Pfam: PF00036; efhand; 2.
DR Pfam: PF00520; Ion_trans; 1.
DR Pfam: PF02815; MIR; 4.
DR Pfam: PF01365; RYDR_ITPR; 2.
DR Pfam: PF02026; RYR; 4.
DR Pfam: PF00622; SPRY; 3.
DR PRINTS: PR00795; RYANODINER.
DR SMART: SM00472; MIR; 4.
DR SMART: SM00449; SPRY; 3.
KW Receptor; Transmembrane; Ionic channel; Calcium channel; Repeat;
KW Phosphorylation; Glycoprotein; Multigene family; Calmodulin-binding.
FT DOMAIN 1 3090 CYTOPLASMIC.
FT TRANSMEM 3091 3110 M' (POTENTIAL).
FT TRANSMEM 3154 3172 M' (POTENTIAL).
FT TRANSMEM 3941 3960 M1 (POTENTIAL).
FT TRANSMEM 3979 3996 M2 (POTENTIAL).
FT TRANSMEM 4234 4257 M3 (POTENTIAL).
FT TRANSMEM 4295 4315 M4 (POTENTIAL).
FT TRANSMEM 4501 4521 M5 (POTENTIAL).
FT TRANSMEM 4580 4602 M6 (POTENTIAL).
FT TRANSMEM 4722 4742 M7 (POTENTIAL).
FT TRANSMEM 4770 4788 M8 (POTENTIAL).
FT TRANSMEM 4812 4829 M9 (POTENTIAL).
FT TRANSMEM 4847 4869 M10 (POTENTIAL).
FT DOMAIN 670 808 SPRY 1.
FT DOMAIN 1098 1221 SPRY 2.
FT DOMAIN 1423 1561 SPRY 3.
FT DOMAIN 853 2926 4 X APPROXIMATE REPEATS.
FT REPEAT 853 966 1.
FT REPEAT 967 1080 2.
FT REPEAT 2693 2811 3.
FT REPEAT 2813 2926 4.
FT BINDING 2619 3016 MODULATOR (POTENTIAL).
FT BINDING 2775 2807 CALMODULIN (POTENTIAL).
FT BINDING 2877 2898 CALMODULIN (POTENTIAL).
FT BINDING 2998 3016 CALMODULIN (POTENTIAL).
FT MOD_RES 2809 2809 PHOSPHORYLATION (BY CAM-KINASE).
FT CARBOHYD 4105 4105 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4796 4796 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 4969 AA; 565060 MW; FF6E0684B974BB4D CRC64;
```

```
Query Match 10.2%; Score 73; DB 1; Length 4969;
Best Local Similarity 25.0%; Pred. No. 3e+02;
Matches 27; Conservative 20; Mismatches 35; Indels 26; Gaps 6;
```

```

Qy      17 YFEPTGPYLMVNTVGDVSKGNELLSPHYVEFPKPGTILTKREIEYV-----EWA 67
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      2681 YMESNYVSMMEKQSSMSEGN--FNPQ---PVDTSNIITPEKLEYFINKYAEHSHDKWS 2734

Qy      68 LDATA----YKEFRVVELDPSAKIE--VTYYDKNKKKKEETKSPITE 108
      : | | | | | : | : | : | : | : | : | : | : | : | : | : | :
Db      2735 MDKLANGWYGEI----YSDSSKIQLPMKPYKLLSEKEKEIYRWPIKE 2778

RESULT  21
YEAM_ECOLI
ID      YEAM_ECOLI          STANDARD;          PRT;          374 AA.
AC      P76253;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, last sequence update)
DT      16-OCT-2001 (Rel. 40, last annotation update)
DE      Putative dioxxygenase alpha subunit yeast (EC 1.14.1.-).
GN      YEAM OR B1802 OR Z2845 OR ECS2511.
OS      Escherichia coli, and
OS      Escherichia coli O157:H7.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562, 83334;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-K12 / MG1655;
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      *The complete genome sequence of Escherichia coli K-12.*;
RL      Science 277:1453-1474(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-K12;
RX      MEDLINE=97251358; PubMed=9097040;
RA      Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA      Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA      Mizobuchi K., Mori H., Mori T., Motomura K., Nakada S., Nakamura Y.,
RA      Nashimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA      Sivasubaram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA      Yamamoto Y., Horluchi T.;
RT      *A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT      corresponding to the 40.1-50.0 min region on the linkage map.*;
RL      DNA Res. 3:379-392(1996).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX      MEDLINE=21074935; PubMed=11206551;
RA      Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA      Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA      Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA      Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA      Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA      Welch R.A., Blattner F.R.;
RT      *Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.*;
RT      Nature 409:529-533(2001).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O157:H7 / RIMD 0509952;
RX      MEDLINE=21156231; PubMed=11258796;
RA      Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA      Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA      Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA      Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT      *Complete genome sequence of enterohaemorrhagic Escherichia coli
RT      O157:H7 and genomic comparison with a laboratory strain K-12.*;
RL      DNA Res. 8:11-22(2001).
CC      -1- COFACTOR: PROBABLY BINDS A 2FE-2S GROUP AND AN IRON ATOM.
CC      -1- SUBUNIT: PROBABLY HETERODIMER OF YEAM AND YEAX.
CC      -1- SIMILARITY: BELONGS TO THE BACTERIAL RING-HYDROXYLATING
CC      DIOXYGENASE ALPHA SUBUNIT FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC
DR EMBL; AE000274; AAC74872.1; -.
DR EMBL; D90823; BAA15597.1; ALT_INIT.
DR EMBL; D90824; BAA15606.1; ALT_INIT.
DR EMBL; AE005403; AAG56791.1; -.
DR EMBL; AP002558; BAB35934.1; -.
DR EcoGene; EG13509; yeast.
DR InterPro; IPR001281; Rleske.
DR InterPro; IPR001663; Ring_hydroxyl_A.
DR Pfam; PF00355; Rleske; 1.
DR PROSITE; PS00570; RING_HYDROXYL_ALPHA; FALSE_NEG.
KM Hypothetical protein; Oxidoreductase; Iron-sulfur; Iron; Dioxygenase;
KW NAD; Complete proteome.
FT METAL 89 89 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 91 91 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 109 109 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 112 112 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 211 211 IRON (BY SIMILARITY).
FT METAL 216 216 IRON (BY SIMILARITY).
SQ SEQUENCE 374 AA; 42561 MW; BB5386AC9585606 CRC64;

```

```

Query Match          10.1%; Score 72.5; DB 1; Length 374;
Best Local Similarity 25.7%; Pred. No. 16;
Matches 37; Conservative 26; Mismatches 38; Indels 43; Gaps 9;

QY      8 KYKKGDDASYFEPTG---PYLMVNTGVDSKGNELLSPHYVEFPKPGTTL-----T 56
      1::| |::| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      251 KFEETDAAFH--GFWLMPCTMLNVTPI--KG--MMTVIY-EFPVDETTLLQNYDIYFT 302
      ::::| |::| | | | | | | | | | | | | | | | | | | | | | | | | |
QY      57 KEKI---EYVEWALDATAAYKEFRVELDPSAKIEVYYDKNKKKETKSPITEKGFV 112
      ::::| |::| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      303 NEELTDEQKSLIEWYRDVFRPEDLRIVE-----SVQGLKSRGYYRGGR I 347
      ::::| |::| | | | | | | | | | | | | | | | | | | | | | | | | |
QY      113 VPD-----LSEHIKNPGFNLTIKV 131
      : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      348 MADSSGSGISEHGIAHFHLLAQV 371

RESULT 22
SURA_BUCAL
ID      SURA_BUCAL          STANDARD;          PRT;          430 AA.
AC      P57240;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Survival protein surA homolog precursor.
GN      SURA OR B0140.
OS      Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
      symbiotic bacterium).
OC      Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX      NCBI_TaxID=118099;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=TOKYO 1998;
RX      MEDLINE=20445173; PubMed=10993077;
RA      Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT      "Genome sequence of the endocellular bacterial symbiont of aphids
      Buchnera sp. APS.";
RL      Nature 407:81-86(2000).
-----
-1- SIMILARITY: BELONGS TO THE PIC/PARVULIN FAMILY OF ROTAMASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```



CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AP001118; BAB12858.1; -.  
DR InterPro: IPR000297; Rotamase.  
DR Pfam: PF00639; Rotamase; 1.  
DR PROSITE: PS01096; PPIC\_PPIASE\_1; FALSE\_NEG.  
DR PROSITE: PS50198; PPIC\_PPIASE\_2; 2.  
KW Isomerase; Rotamase; Repeat; Signal; Complete proteome.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 430 SURVIVAL PROTEIN SURA HOMOLOG.  
FT DOMAIN 179 277 PPIC 1.  
FT DOMAIN 286 386 PPIC 2.  
SQ SEQUENCE 430 AA; 50737 MW; CAAC425B3FBA5788 CRC64;  
  
Query Match 10.1%; Score 72.5; DB 1; Length 430;  
Best Local Similarity 22.2%; Pred. No. 18;  
Matches 39; Conservative 28; Mismatches 44; Indels 65; Gaps 12;  
  
QY 9 YKGGDASYFEPTGPLY-----MVNVGVDSKGNELLSPHYVEF-PIKPGTTLT---K 57  
Db 253 FKKG-----QILGPIVGDKGLYILKVNDRHKKENIVTEFYMOHCLIKPSVILTNTAK 306  
QY 58 EKIEYVEMALDATAKKEFR--VVELDPSAK-IEVTYYDKNKK-----KETKSEPTTE 108  
Db 307 KKI-----FNIEYENIKKGIYTFDDAVKNSLSDYYSSNKKGDLGWSKESLGFDLAK 357  
QY 109 KGFVVPDLSE-----HI-----KNPGFNLI--TKVIEK 135  
Db 358 K-FLLDKNEISEPYKSNMGWHIFKILDRQVDAYFKLKKNAFNIVLNOKIISEK 412  
  
RESULT 23  
ID IF38\_MEDTR STANDARD; PRT; 935 AA.  
AC O9XHM1;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Probable eukaryotic translation initiation factor 3 subunit 8 (eIF3  
DE p110).  
GN AM3-1.  
OS Medicago truncatula (Barrel medic).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
OX NCBI\_TaxID=3880;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A17;  
RX MEDLINE=99164949; Pubmed=10065555;  
RA van Buuren M.L., Maldonado-Mendoza I.E., Trieu A.T., Blaylock L.A.,  
RA Harrison M.J.;  
RT "Novel genes induced during an arbuscular mycorrhizal (AM) symbiosis  
RT formed between Medicago truncatula and Glomus versiforme.";  
RL Mol. Plant Microbe Interact. 12:171-181(1999).  
CC -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF  
CC METHIONYL-TRNAI AND MRNA (BY SIMILARITY).  
CC -1- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE EIF3S8 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: AF106930; AAD39891.1; -.  
DR InterPro: IPR000717; PCT.  
DR Pfam: PF01399; PCT; 1.  
DR SMART: SM00088; PINT; 1.  
KW Initiation factor; Protein biosynthesis.  
SQ SEQUENCE 935 AA; 105069 MW; 199F5FF898EB1F CRC64;  
  
Query Match 10.1%; Score 72.5; DB 1; Length 935;  
Best Local Similarity 23.0%; Pred. No. 46;  
Matches 37; Conservative 31; Mismatches 50; Indels 43; Gaps 11;  
  
QY 1 SSSFDKGGKKGGDASYFEPTGPLYMNVWG-----VDS---KGNELLSPH---YVE 46  
Db 347 SVELDENETKKGDYNN-----GP---INWGNLVAFLKIDAEFFKSLQCIDPHTRYVE 398  
QY 47 -FPIKPGTTLTKEKIEYVEMALDATAKKEFRVVELDPSAKIEVTYYDKNKKKEETKSF- 104  
Db 399 RLKDEPQFVLAQNVQETLE-----SIGDFKASSKVALKRVELIYKPKHVEYEAIRKLA 452  
QY 105 PITEKGFVVPDLSEHIKNP-GFN-----LITKRVIEKK 136  
Db 453 EMIVEG-----DNGEMSEEPKGFEDTRIPAPFVYTLVLARK 489

RESULT 24  
ID SYI\_CHLPN STANDARD; PRT; 1043 AA.  
AC Q9Z972; Q9JQK2; Q9K221;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Isolation of Chlamydia pneumoniae (EC 6.1.1.5) (Isolation of Chlamydia  
DE pneumoniae).  
GN IL6S OR CPN0109 OR CP0665.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiales.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; Pubmed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; Pubmed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CWL029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
CC -1- CATALYTIC ACTIVITY: ATP + L-isoleucine + tRNA(Ile) = AMP +  
CC diphosphate + L-isoleucyl-tRNA(Ile).  
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -----

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001597; AAD18262.1; -.
DR EMBL; AE002224; AAF38477.1; -.
DR EMBL; AP002545; BAA98320.1; -.
DR HSSP; P56690; 1ILE.
DR TIGR; CP0665; -.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002301; tRNA-synt_1le.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00984; TRNASYNTHILE.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminocacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; zinc; Complete proteome.
FT SITE 48 58 "HIGH" REGION.
FT SITE 591 595 "KMSKS" REGION.
FT BINDING 594 594 ATP (BY SIMILARITY).
FT CONFLICT 384 384 H -> Y (IN REF. 2).
SQ SEQUENCE 1043 AA; 120286 MW; 08ABB8B9364E640C CRC64;

Query Match 10.1%; Score 72.5; DB 1; Length 1043;
Best Local Similarity 20.8%; Pred. No. 53;
Matches 26; Conservative 24; Mismatches 42; Indels 33; Gaps 5;

QY 24 YLMVNTGVDSKGNELISPHYVEFPKPGTTLTKKEIYVE-----W-----AL 68
DB 680 WILSNLYSVGKVRRESMSQYHLNFAVEPFVTIDLTFWYIRRCRRRFEAEPTDPRRA 739
QY 69 DATAKKEFRVELDPSAKIEVTV-----YDKNKKKEETKS-----FPITERGFVVPD 115
DB 740 FSTLYEVLTVF-----CKVIAFVFPFLAEDIYQKTKLEKEPESVHLCDFPQVEMDKILPD 794
QY 116 LSEHI 120
DB 795 LEKRM 799

RESULT 25
YQ30_BACAN
ID YQ30_BACAN STANDARD; PRT; 354 AA.
AC Q9RND2;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein pXO2-30.
GN pXO2-30.
OS Bacillus anthracis.
OC Plasmid pXO2.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RA Okinaka R.T., Cloud K., Hampton O., Hill R.K., Keim P., Lamke G.,
RA Kumano S., Manter D., Martinez Y., Svensson R., Tatum L.R.,
RA Brown A.E., Jackson P.J.;
RA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC      or send an email to license@lsb-sib.ch).
CC      -----
DR      EMBL; AF188935; AAF13635.1; "
KW      Hypothetical protein; Plasmid; Transmembrane.
FT      TRANSMEM      8      28      POTENTIAL.
SQ      SEQUENCE      354 AA; 39491 MW; F4E5569E00FB0F7A CRC64;

Query Match
Best Local Similarity 26.3%; Score 72; DB 1; Length 354;
Matches 36; Conservative 17; Mismatches 58; Indels 26; Gaps 5;

OY      10 KKGDASYFEP-----TGPYLMVNTGVDSKGNELISPHY--VEFPKPGTTLTK 57
          ||| | || | | | | | | | | | | | | | | | | | | | | | | | | |
DQ      69 KKGSDDLDFEAFIKPSDIDVGTEPFDKSNENAGNDSSPNNGIVRTFDVTYPLK--VTINP 126
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      58 EKIEYVEMALDATAYKEFRVELDPSAKIEVTVYDKNKKKEETKSPITEKGFVVDLS 117
          : | : : | | | | | : : || | | | | | | | | | | | | | | | |
DQ      127 KKHDKLKIDIKLKITGTLENGITNKRNVNAKFAV-----GKED-----LEKGIVSFDM 174
          || | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      118 EHIKNPGFNLTQVIE 134
          : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ      175 YTLKETGNSLMIPVTE 191

RESULT 26
HISX_METUA STANDARD; PRT; 429 AA.
AC      Q58851;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Histidinol dehydrogenase (EC 1.1.1.23) (HDH).
GN      HISD OR MJ1456.
OS      Methanococcus jannaschli.
OC      Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC      Methanococcus.
OX      NCBI_TaxID=2190;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX      MEDLINE=96337999; PubMed=8688087;
RA      Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA      Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA      Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA      Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA      Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA      Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA      Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA      Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT      "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT      jannaschli.";
RL      Science 273:1058-1073(1996).
CC      -1- FUNCTION: THIS PROTEIN IS CONSIDERED AS A BIFUNCTIONAL ENZYME,
CC      POSSESSING TWO ACTIVE SITES, ONE AN ALCOHOL DEHYDROGENASE AND
CC      THE OTHER AN ALDEHYDE DEHYDROGENASE (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: L-histidinol + 2 NAD(+) + H(2)O = L-histidine
CC      + 2 NADH.
CC      -1- PATHWAY: TENTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
CC      -1- SIMILARITY: TO OTHER PROKARYOTIC, FUNGAL AND PLANTS HDH.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@lsb-sib.ch).
CC      -----
DR      EMBL; U67586; AAB99465.1; -
DR      TIGR; MJ1456; -
DR      InterPro; IPR001692; Histidinol_dh.
DR      Pfam; PF00815; Histidinol_dh; 1.

```



```
DR PRINTS; PR00083; HOLDHGRNASE.
DR PRODOM; PD002680; Histidinol_dh; 1.
DR PROSITE; PS00611; HISOL_DEHYDROGENASE; 1.
KW Histidine biosynthesis; Multifunctional enzyme; Oxidoreductase; NAD;
KW Complete proteome.
SQ SEQUENCE 429 AA; 47432 MW; 5E4DF4B17699EDB8 CRC64;

Query Match 10.1%; Score 72; DB 1; Length 429;
Best Local Similarity 25.9%; Pred. No. 20;
Matches 28; Conservative 21; Mismatches 51; Indels 8; Gaps 4;

QY 18 FEPTGPLYMNVTVGVDSKGNELSPHYVEFPKPGTTLTKETI-EYVEMALDATAYKEF 76
DB 29 FEELPTVMEILKDVKEKGDEALKYTKKF--DGEVEIDFKYDEIEEAYNSVDYK-- 83

QY 77 RVVELDPSAKIEVTYYVDKNKKKEETKSPITEKGFVVPDLSEHIKNPG 124
DB 84 -VVEAIERAKENYFFHK-KOMEQIKDLNVENNIGIILGVVRAIEKVG 129

RESULT 27
APP2_MOUSE STANDARD; PRT; 695 AA.
AC 006335;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Amyloid-like protein 2 precursor (CDEI-box binding protein) (CDEBP).
GN APLP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RA von der Kammer H.;
RL Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-246 FROM N.A.
RX MEDLINE=94032480; PubMed=8218408;
RA Hanes J., von der Kammer H., Kristjansson G.I., Scheit K.H.;
RT "The complete cDNA coding sequence for the mouse CDEI binding
RT protein.";
RL Biochim. Biophys. Acta 1216:154-156(1993).
RN [3]
RP SEQUENCE OF 185-695 FROM N.A.
RC STRAIN=BALB/C; TISSUE=Heart;
RX MEDLINE=93129193; PubMed=1482349;
RA Vidal F., Blangy A., Rassoulzadegan M., Cuzin F.;
RT "A murine sequence-specific DNA binding protein shows extensive local
RT similarities to the amyloid precursor protein.";
RL Biochem. Biophys. Res. Commun. 189:1336-1341(1992).
RN [4]
RP SEQUENCE OF 1-35 FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=96029629; PubMed=7592716;
RA von Koch C.S., Lahiri D.K., Mammen A.L., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Sisodia S.S.;
RT "The mouse APLP2 gene. Chromosomal localization and promoter
RT characterization.";
RL J. Biol. Chem. 270:25475-25480(1995).
CC -1- FUNCTION: BINDS TO THE DNA 5'-GTCACATG-3' (CDEI BOX) WHICH PLAYS
CC AN IMPORTANT ROLE IN THE EARLY DEVELOPMENT OF EMBRYOS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN AND NUCLEAR
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22592; CAA80306.1; -
DR EMBL; M97216; AAA20039.1; -
DR EMBL; U34291; AAC52318.1; -
DR PIR; JC1404; JC1404.
DR HSSP; P05067; 1MWP.
DR MGD; MGI:88047; APLP2.
DR InterPro; IPR001868; A4_APP.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR SMART; SM00006; A4_EXTRA; 1.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
KW Transmembrane; DNA-binding; Signal; Nuclear protein.
FT SIGNAL 1 29
FT CHAIN 30 695
FT DOMAIN 30 624
FT TRANSMEM 625 648
FT DOMAIN 649 695
FT DOMAIN 218 294
FT DOMAIN 218 231
FT DOMAIN 256 266
FT CARBOHYD 485 485
FT CONFLICT 185 189
SQ SEQUENCE 695 AA; 78944 MW; BBF4B95AAB2A0311 CRC64;

Query Match 10.1%; Score 72; DB 1; Length 695;
Best Local Similarity 23.8%; Pred. No. 36;
Matches 29; Conservative 21; Mismatches 44; Indels 28; Gaps 6;

QY 11 KDDASYFEPTGPLYMNVTVGVDSKGNELSPHYVEFPK-PGTTLTKEIEYVEMALD 69
DB 279 KGDDYNEENPTEP-----SSEGT--ISDKREIVHDVKVPPTPLPTNDYDYLLET SAD 327

QY 70 ATAYKEFRVELDPSAKIEVTYYVDKNK---KKE-----ETKSFPTTEKGFVVPDLSEH 119
DB 328 DNEHARFQ-----KAKEQLEIRHRNRMDRVKKKEAEALQAKNLPKTERGTLLIQHFOAM 381

QY 120 IK 121
DB 382 VK 383

RESULT 28
KFMS_RAT STANDARD; PRT; 978 AA.
ID KFMS_RAT
AC 000495;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Macrophage colony stimulating factor 1 receptor precursor (CSF-1-R)
DE (EC 2.7.1.112) (Fms proto-oncogene) (c-fms).
GN CSF1R OR CSFMR OR FMS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=Skeletal muscle;
RX MEDLINE=93001225; PubMed=1389227;
RA Borycki A.G., Guilleter M., Leibovitch M.P., Leibovitch S.A.;
RT "Molecular cloning of CSF-1 receptor from rat myoblasts. Sequence
RT analysis and regulation during myogenesis.";
RL Growth Factors 6:209-218(1992).
CC -1- FUNCTION: THIS PROTEIN IS THE RECEPTOR FOR CSF-1, IT IS A PROTEIN
CC TYROSINE-KINASE TRANSMEMBRANE RECEPTOR.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
```

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-  
 CC PROTEIN KINASES.  
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X61479; CAA43706.1; -.  
 DR PIR; S16385; S16385.  
 DR HSSP; P11362; IFGK.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003598; Ig\_C2.  
 DR InterPro; IPR003600; Ig\_Like.  
 DR InterPro; IPR01824; Receptor\_tyr\_kin\_III.  
 DR InterPro; IPR01245; Tyr\_pkinase.  
 DR Pfam; PF00047; Ig; 4.  
 DR Pfam; PF00069; pkinase; 2.  
 DR SMART; SM00410; Ig\_Like; 3.  
 DR SMART; SM00408; IgC2; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; RECEPTOR\_TYR\_KIN\_III; 1.  
 DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Proto-oncogene; Tyrosine-protein kinase; Receptor; Transmembrane;  
 KW transferase; Glycoprotein; Phosphorylation; ATP-binding; Signal;  
 KW Immunoglobulin domain; Repeat.  
 FT SIGNAL 1 19  
 FT CHAIN 20 978  
 FT DOMAIN 20 511  
 FT TRANSMEM 512 536  
 FT DOMAIN 537 978  
 FT DOMAIN 24 104  
 FT DOMAIN 107 197  
 FT DOMAIN 204 298  
 FT DOMAIN 299 397  
 FT DOMAIN 398 503  
 FT DOMAIN 580 914  
 FT NP\_BIND 586 594  
 FT BINDING 614 614  
 FT ACT\_SITE 776 776  
 FT DISULFID 42 84  
 FT DISULFID 127 177  
 FT DISULFID 224 278  
 FT DISULFID 417 483  
 FT MOD\_RES 697 697  
 FT MOD\_RES 706 706  
 FT MOD\_RES 807 807  
 FT CARBOHYD 45 45  
 FT CARBOHYD 73 73  
 FT CARBOHYD 302 302  
 FT CARBOHYD 335 335  
 FT CARBOHYD 389 389  
 FT CARBOHYD 410 410  
 FT CARBOHYD 449 449  
 FT CARBOHYD 478 478  
 FT CARBOHYD 491 491  
 SO SEQUENCE 978 AA; 109264 MW; 0A68456EF56BC7E3 CRC64;

Query Match 10.1%; Score 72; DB 1; Length 978;  
 Best Local Similarity 23.2%; Pred. No. 54;

Matches 35; Conservative 20; Mismatches 44; Indels 52; Gaps 6;  
 QY 12 GDDASYEPTGYLMVN-----VTGVDKGNELLSPHYVEFPKPGTLTKK--- 59  
 Db 17 GGGAPVIEPSGPELVPEGTTLRCVSNGSVWDGPISPYWTLDPESGSLTTRNATF 76  
 QY 60 -----IEYVE-----WALDATAKFRVVE-----LDPS 84  
 Db 77 KMTGYRCTELEDPMGSTTILYVKDPAHSWNLLA---QEVTVGEQAVLPCLITDPA 133  
 QY 85 AKIEVYYDDKNKK--EETKSPITEKGFV 113  
 Db 134 LKDSVSLMREGGRQVLRKRYTFESAMRGFTI 164  
 RESULT 29  
 YE20\_METUA STANDARD; PRT; 1102 AA.  
 AC Q58815;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ1420 [Contains: Mja gfp6 Intein].  
 GN MJ1420.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
 OC Methanococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reisch C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii";  
 RL Science 273:1058-1073(1996).  
 CC -1- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES  
 CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).  
 CC -1- SIMILARITY: IN THE C-TERMINAL, TO GLUCOSAMINE--FRUCTOSE-6-  
 CC PHOSPHATE AMINOTRANSFERASES (GFAT).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U67582; AAB99430.1; -.  
 DR HSSP; P17169; LMOQ.  
 DR MEROPS; C44.971; -.  
 DR TIGR; MJ1420; -.  
 DR InterPro; IPR000583; GATase\_2.  
 DR InterPro; IPR001387; HTH\_3.  
 DR InterPro; IPR003586; Hintc.  
 DR InterPro; IPR003587; Hintn.  
 DR InterPro; IPR002203; Intein.  
 DR InterPro; IPR004042; Intein\_endonuc.  
 DR InterPro; IPR001347; SIS.  
 DR Pfam; PF00310; GATase\_2; 2.  
 DR Pfam; PF01381; HTH\_3; 1.  
 DR Pfam; PF01380; SIS; 2.  
 DR SMART; SM00530; HTH\_XRE; 1.  
 DR SMART; SM00305; Hintc; 1.

```
DR SMART; SMO0306; HintN; 1.  
DR PROSITE; PS00443; GATASE_TYPE_II; 1.  
DR PROSITE; PS00881; PROTEIN_SPLICING; FALSE_NEG.  
KW Hypothetical protein; Autocatalytic cleavage; Protein splicing;  
Complete proteome.  
FT CHAIN 1 74 HYPOTHETICAL PROTEIN MJ1420, 1ST PART  
FT FT CHAIN 75 573 (POTENTIAL).  
FT FT CHAIN 574 1102 MJA GF6P INTEIN (POTENTIAL).  
FT SEQUENCE 1102 AA; 125907 MW; 9CIEFA8BD890CBAA1 CRC64;  
  
Query Match 10.1%; Score 72; DB 1; Length 1102;  
Best Local Similarity 25.5%; Pred. No. 62;  
Matches 41; Conservative 21; Mismatches 51; Indels 48; Gaps 11;  
  
QY 6 KGVKKKGDDASY-FEPTGYLM-----VNVTVGDSK-GNELLSPHYVEFPPIKGTTLTKK 59  
DB 321 EGNIKKGDNYYILEINSKYLLIDWERENIPLENKTGNE-RTPEFV-----FRLNNDL 372  
QY 60 IEXYEWALDATAY--KEFRVELDPSAKI---EVTY-----YDKNKKEET--- 101  
DB 373 VASILRGIFDAEGYIRAEAKQIGIMTSKCFIKEIQFLLRFGILASYSKIKRKEENMNN 432  
QY 102 -----KSPITEK--GEVVPDLSEHI-----KNPGFN 126  
DB 433 THKLISDKKSFELEFKYIGFTAKDKMEKLEAILNMKGGLN 473  
  
RESULT 30  
CENE_HUMAN STANDARD; PRT; 2663 AA.  
ID CENE_HUMAN  
AC Q02224;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Centromeric protein E (CENP-E protein).  
CN CENPE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
ON NCBI_TaxID=9606;  
RX [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93024922; PubMed=1406971;  
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;  
RT "CENP-E is a putative kinetochore motor that accumulates just before  
mitosis.";  
RL Nature 359:536-539(1992).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=95196755; PubMed=7889940;  
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;  
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed  
microtubule motor.";  
RL EMBO J. 14:918-926(1995).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=98437347; PubMed=9763420;  
RA Chan G.K.T., Schaar B.T., Yen T.J.;  
RT "Characterization of the kinetochore binding domain of CENP-E reveals  
interactions with the kinetochore proteins CENP-F and hBUBR1.";  
RL J. Cell Biol. 143:49-63(1998).  
CC -|- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE  
KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE  
OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT  
AND/OR SPINDLE ELONGATION.  
CC -|- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.  
CC -|- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING  
CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS  
QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.  
CC -|- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
```

```

-----CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----CC
DR EMBL; Z15005; CAA78727.1; -.
DR PIR; S28261; S28261.
DR HSSP; P17119; 3KAR.
DR MIM; I17143; -.
DR InterPro; IPR001752; kinesin.
DR Pfam; PF00225; kinesin_1.
DR PRINTS; PRO0380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT FTDOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 2663 AA; 312087 MW; CEFC13880C8C8CB8 CRC64;

Query Match 10.1%; Score 72; DB 1; Length 2663;
Best Local Similarity 28.7%; Pred. No. 1.8e+02;
Matches 25; Conservative 16; Mismatches 26; Indels 20; Gaps 4;

QY 27 VNTGVDSKGNELSPHYVEFPKPGTLTKETI--EYYEWALDATAKFRRVEL-- 81
||| |::| :|| | |||| | : |::| ::::
Db 2321 VNPTIQDNK-----NPHVT-----SRATQLTEKIRELENSLHKAESAMHKSKIITKOK 2371

QY 82 -----DPSAKIEVTTYDKNKKEETK 102
| ||:: :: || |::|

Db 2372 ELEVTNDIIIAKLQAKVHESNKCLEKTK 2398

RESULT 31
YF61_METJA STANDARD; PRT; 553 AA.
ID YF61_METJA
AC Q58956;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1561.
GN MJ1561.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
SC MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.G., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weldman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kalne B.P., Borodovsky M.,
RA Kleen H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO M.JANNASCHII MJ0795 AND MJ1506.
-----CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U67596; AAB99582.1; -.  
DR TIGR; MJ1561; -.  
KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRANSMEM 6 26 POTENTIAL.  
FT TRANSMEM 524 544 POTENTIAL.  
SQ SEQUENCE 553 AA; 61713 MW; 86CA1E67C5A0AA01 CRC64;

Query Match 10.0%; Score 71.5; DB 1; Length 553;  
Best Local Similarity 23.1%; Pred. No. 31;  
Matches 36; Conservative 24; Mismatches 43; Indels 53; Gaps 9;

OY 7 GKXKKG-----DASYEPTGPLYMNVTVGVDSKG-----NELLSPH 43

Db 283 GAYVKGDIILGISVNTDPKEIKPGTYVRIDVT-ITNGHAEAKDYKLIKLTNKPFDK 341

OY 44 YVEFPIK-----PGTTLTKEKIEYVEMALDATAKKEFRVELDPSAKIEVYYDKNK 97

Db 342 WSNCKIKDVGNLPGVSKT---VSFYVDVDRYASA-KHYKL-----PIEISYLDTANN 390

OY 98 KEETKSPITEKGFVVPDLSEHIK-NPGFNLTITKV 132

Db 391 KYKTEK-----IDIVKPKPLFEITKEV 415

RESULT 32

POTA\_MYCPN

ID POTA\_MYCPN STANDARD; PRT; 560 AA.

AC P75059;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Spermidine/putrescine transport ATP-binding protein potA homolog.

GN POTA OR MPN055 OR MP099.

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscites;

OC Mycoplasmataceae; Mycoplasma.

OX NCBI\_TaxID=2104;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29342 / M129;

RX MEDLINE=97105885; PubMed=8948633;

RA Himmelfreih R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,

RA Herrmann R.;

RT \*Complete sequence analysis of the genome of the bacterium Mycoplasma

RT pneumoniae.\*;

RT Nucleic Acids Res. 24:4420-4449(1996).

CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM

CC FOR SPERMIDINE/PUTRESCINE. PROBABLY RESPONSIBLE FOR ENERGY

CC COUPLING TO THE TRANSPORT SYSTEM (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Membrane-associated (Potential).

CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AE000011; AAB95747.1; -.

DR HSSP; P13569; INBD.

DR InterPro; IPR003593; AAA.

DR InterPro; IPR003439; ABC\_transportr.

DR InterPro; IPR001687; ATP\_GTP\_A.

DR Pfam; PF00005; ABC\_tran; 1.

DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
KW Transport; ATP-binding; Membrane; Complete proteome.  
FT NP\_BIND 40 47 ATP (POTENTIAL).  
SQ SEQUENCE 560 AA; 65131 MW; 8993CAE61A28723D CRC64;

Query Match 10.0%; Score 71.5; DB 1; Length 560;  
Best Local Similarity 22.5%; Pred. No. 31;  
Matches 39; Conservative 22; Mismatches 51; Indels 61; Gaps 9;

OY 1 SSSFDKG-----KXKGDASYFEPTG-----PYLMVNTGVDSK 35

Db 13 NKTFDGFSVSRDINLKIKGFEITILGPSGCGKTTTLRLLAGFEDPTYGKIKVNGLDIK 72

OY 36 GNELLSPHYVEFPI--KPGTTLTKEKIEYVEMAL--DATAKKEF-----RVVELDPSA 85

Db 73 -----DLPIHKRPFAFV-----FQDYALFSLTVYKNIAAYGLKAMYTKLDPID 115

OY 86 KIEVYYDK--NKKKEETKSPITEKGFVVPDLSEHIKNGFNLTITKVIEKK 136

Db 116 KIVEQYHOSLLDKQHLRHKRIERLEKSNANPOLLEOLKO-----TVVVOQR 161

RESULT 33

Y366\_MYCGE

ID Y366\_MYCGE STANDARD; PRT; 667 AA.

AC P47606;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein MG366.

GN MG366.

OS Mycoplasma genitalium.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscites;

OC Mycoplasmataceae; Mycoplasma.

OX NCBI\_TaxID=2097;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 33530 / G-37;

RX MEDLINE=96026346; PubMed=7569993;

RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,

RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,

RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,

RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,

RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,

RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;

RT "The minimal gene complement of Mycoplasma genitalium.";

RL Science 270:397-403(1995).

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U39718; AAC71593.1; -.

DR TIGR; MG366; -.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 667 AA; 78117 MW; 450389698D0690D8 CRC64;

Query Match 10.0%; Score 71.5; DB 1; Length 667;  
Best Local Similarity 26.5%; Pred. No. 38;  
Matches 27; Conservative 16; Mismatches 40; Indels 19; Gaps 4;

OY 46 EFPKPGTTLTKEKIEYVEMALDATAKKE-----FRVELDPSAKIEVYYDKN 95

Db 162 EFVVIKATNTTKRK--YFLEIYDFVLFKKIGKYLKLANFYCTVKEYLQNKNNVSFFLNT 219

OY 96 KKKETKSPITEKGFVVPDLSEHIKNGFNLTITKV--IEKK 136

Db 220 EIKTSKNSFSLSKE-----KDYFKNKPENHPEKIAYIHKK 255

RESULT 34  
FREL\_YEAST STANDARD; PRT; 686 AA.

ID FREL\_YEAST  
AC P32791;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ferric reductase transmembrane component 1 precursor (EC 1.6.99.13)  
DE (Ferric-chelate reductase 1).  
GN FREL OR YLR214W OR L8167.2.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=F113;  
RX MEDLINE=92237270; PubMed=1570306;  
RA Dancis A., Roman D.G., Anderson G.J., Hinnebusch A.G., Klausner R.D.;  
RT "Ferric reductase of Saccharomyces cerevisiae: molecular  
RT characterization, role in iron uptake, and transcriptional control by  
RT iron."  
RL Proc. Natl. Acad. Sci. U.S.A. 89:3869-3873(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
RA Johnston L., Langston Y., Latreille P., Mardis E., Menezes S.,  
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,  
RA Talc S., Trevasakis E., Vaudin M., Vignati D., Wilcox L., Wilson R.,  
RA Wohldman P., Waterston R.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=93057491; PubMed=1431884;  
RA Anderson G.J., Lesuisse E., Dancis A., Roman D.G., Labbe P.,  
RA Klausner R.D.;  
RT "Ferric iron reduction and iron assimilation in Saccharomyces  
RT cerevisiae";  
RL J. Inorg. Biochem. 47:249-255(1992).  
CC -1- FUNCTION: REDUCTASE ACTIVITY THAT ACTS ON FERRIC IRON CHELATES  
CC EXTERNAL TO THE CELL. PLAYS A ROLE IN IRON UPTAKE. MAY  
CC PARTICIPATES IN THE TRANSPORT OF ELECTRONS FROM CYTOPLASM TO AN  
CC EXTRACELLULAR SUBSTRATE (FERRIC ION) VIA FAD AND HEME  
CC INTERMEDIATES. MAY ALSO PARTICIPATE IN CU(II) REDUCTION AND CU(I)  
CC UPTAKE.  
CC -1- CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).  
CC -1- COFACTOR: FAD (PROBABLE).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- INDUCTION: BY IRON DEPRIVATION. REPRESSED BY IRON UPTAKE.  
CC -1- SIMILARITY: BELONGS TO THE FRE / CYBB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M86908; AAA34608.1; -  
DR EMBL; U14913; AAB67424.1; -  
DR PIR; S30075; S30075.  
DR PIR; S48565; S48565.  
DR SGD; S0004204; FREL.  
DR InterPro; IPR002916; Ferric\_reduct.  
DR Pfam; PF01794; Ferric\_reduct; 1.  
CC

KW Oxidoreductase; Electron transport; Transmembrane; Iron transport;  
KW FAD; NAD; Copper; Glycoprotein; Signal; Multigene family.  
FT SIGNAL 1 22  
FT CHAIN 23 686  
FT NP\_BIND 462 468  
FT NP\_BIND 532 540  
FT TRANSMEM 147 169  
FT TRANSMEM 216 236  
FT TRANSMEM 258 277  
FT TRANSMEM 296 316  
FT TRANSMEM 329 348  
FT TRANSMEM 369 397  
FT TRANSMEM 529 550  
FT CARBOHYD 69 69  
FT CARBOHYD 100 100  
FT CARBOHYD 124 124  
FT CARBOHYD 209 209  
FT CARBOHYD 503 503  
FT CARBOHYD 646 646  
SO SEQUENCE 686 AA; 78853 MW; 7F6BB3B93A95D6A3 CRC64;

Query Match 10.0%; Score 71.5; DB 1; Length 686;  
Best Local Similarity 23.0%; Pred. No. 40;  
Matches 29; Conservative 24; Mismatches 44; Indels 29; Gaps 6;

OY 22 GPY-----LMVNTGVDS-KGNELSPHYVEFPKPGTTLREKIEYVWALDAT 71  
Db 514 GPYGVTPHIAKLKRNLVGAAGLVAIAHPHFVECLRLPST---DQLHKFYIWNVL 569  
OY 72 AYKEFRVVEL---DPSAKIEVTV---YDKNKKKEETKSPITEKGFVY-----PDL 116  
Db 570 SHLKWFEENQWLKEKSESVIYTGSSVEDTNSDESTKGFDDKESEITVECLNKRPLD 629  
OY 117 SEHIKN 122  
Db 630 KELVRS 635

RESULT 35  
CYSA\_SYNP7 STANDARD; PRT; 344 AA.  
ID CYSA\_SYNP7  
AC P14788;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Sulfate transport ATP-binding protein cysA.  
GN CYSA.  
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=1140;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89184544; PubMed=2538823;  
RA Green L.S., Laudenbach D.E., Grossman A.R.;  
RT "A region of a cyanobacterial genome required for sulfate transport."  
RL Proc. Natl. Acad. Sci. U.S.A. 86:1949-1953(1989).  
RN [2]  
RP SEQUENCE OF 1-11 FROM N.A.  
RX MEDLINE=91210162; PubMed=1708375;  
RA Laudenbach D.E., Grossman A.R.;  
RT "Characterization and mutagenesis of sulfur-regulated genes in a  
RT cyanobacterium: evidence for function in sulfate transport."  
RL J. Bacteriol. 173:2739-2750(1991).  
CC -1- FUNCTION: INVOLVED IN THE TRANSPORT OF MULTIPLE SULFUR-CONTAINING  
CC COMPOUNDS, INCLUDING SULFATE AND THIOSULFATE.  
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated.  
CC -1- INDUCTION: BY SULFUR DEPRIVATION.  
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -





```

RC STRAIN=LT2;
RX MEDLINE=91260454; Pubmed=1710759;
RA Jiang X.-M., Neal B., Santiago F., Lee S.J., Romana L.K., Reeves P.R.;
RT "Structure and sequence of the rfb (O antigen) gene cluster of
RL Salmonella serovar typhimurium (strain LT2).";
RN Mol. Microbiol. 5:695-713(1991).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; Pubmed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RT Courtney L., Porwollik S., Alt J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RL LT2.";
RL Nature 413:852-856(2001).
CC -1- FUNCTION: INVOLVED IN GDP-MANNOSE BIOSYNTHESIS WHICH SERVES AS
CC THE ACTIVATED SUGAR NUCLEOTIDE PRECURSOR FOR MANNOSE RESIDUES
CC IN CELL SURFACE POLYSACCHARIDES. THIS ENZYME PARTICIPATES IN
CC SYNTHESIS OF THE LPS GROUP B O ANTIGEN.
CC -1- CATALYTIC ACTIVITY: GDP + D-mannose 1-phosphate = phosphate + GDP-
CC mannose.
CC -1- PATHWAY: GDP-MANNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF MANNOSE-6-PHOSPHATE ISOMERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC CC
CC EMBL: X56793; CAA40128.1; -.
CC DR EMBL: AE008792; AAL20988.1; -.
CC DR PIR: S15312; S15312.
CC DR StyGene: SG10350; rfbM.
CC DR InterPro: IPR001538; Mannosep_isomer.
CC DR InterPro: IPR001825; NTP_transferase.
CC DR Pfam: PF01050; Mannosep_isomer; 1.
CC DR Pfam: PF00483; NTP_transferase; 1.
CC DR ProDom: PD002664; Mannosep_isomer; 1.
CC DR Lipopolysaccharide biosynthesis; Transferase; Kinase;
CC Complete proteome.
CC KW
CC SEQUENCE 479 AA; 54046 MW; FE9E84D5CEE1F118 CRC64;

Query Match 9.9%; Score 71; DB 1; Length 479;
Best Local Similarity 22.9%; Pred. No. 29;
Matches 30; Conservative 21; Mismatches 48; Indels 32; Gaps 7;

OY 5 DKCKYKKGDDASYFEPTCPYLWVWTCVDSKGNELLSPHYVEFPPIKPGTTLTKEKIEYV 64
   1: :||: : | | | | | :|| | | :
Db 352 DRNEYKHQHEV--FRPWGKY---NV--IDSGKNYLVRC-----ITVKPGEKFVAQMHHR 400
   1: :||: : | | | | | :|| | | :
OY 65 E-WALDATAIYKEFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVDLSEHIKNP 123
   1: | | : | :||: : | | | | | | | | : ||
Db 401 EHW-----IVLSGTARV-----TKGEQTVMVSENESTFIPNTIHAIENP 440
   1: | | : | :||: : | | | | | | | | : ||

OY 124 GFNLITKRVIE 134
   1: : | : | :
Db 441 GMTPLKLEIQ 451

RESULT 38
PUR9_BACSU
ID PUR9_BACSU STANDARD: PRT; 512 AA.
AC P12048;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)

```

```

DE 16-OCT-2001 (Rel. 40, last annotation update)
DE Bifunctional purine biosynthesis protein purH [includes:
DE phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)
DE (AICAR transformylase); IMP cyclohydrolase (EC 3.5.4.10) (Inosinase)
DE (IMP synthetase) (ATIC)].
GN purH OR purHJ.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87250425; PubMed=3036807;
RA Ebbole D.J., Zalkin H.;
RT "Cloning and characterization of a 12-gene cluster from Bacillus
RT subtilis encoding nine enzymes for de novo purine nucleotide
RT synthesis.";
RL J. Biol. Chem. 262:8274-8287(1987).
RN [2]
RP SEQUENCE OF 432-512 FROM N.A.
RX STRAIN=168;
RC MEDLINE=97124186; PubMed=8969499;
RA Borris R., Porwollik S., Schroeter R.;
RT "The 52 degrees-55 degrees segment of the Bacillus subtilis
RT chromosome: a region devoted to purine uptake and metabolism, and
RT containing the genes cotA, gabP and guaA and the pur gene cluster
RT within a 34960 bp nucleotide sequence.";
RL Microbiology 142:3027-3031(1996).
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5'-
CC phosphoribosyl-5-amino-4-imidazolecarboxamide = tetrahydrofolate +
CC 5'-phosphoribosyl-5-formamido-4-imidazolecarboxamide.
CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-
CC phosphoribosyl)imidazole-4-carboxamide.
CC -1- PATHWAY: NINTH AND TENTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC -1- DOMAIN: THE IMP CYCLOHYDROLASE ACTIVITY RESIDES IN THE N-TERMINAL
CC REGION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE purH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: J02732; AAA22683.1; -.
CC DR EMBL: AF011544; AAB72185.1; -.
CC DR EMBL: Z99107; CAB12472.1; -.
CC DR PIR: A29183; DTBSPH.
CC DR Subtilist; BG10710; purH.
CC DR InterPro: IPR002695; AICARFT_IMPCHas.
CC DR InterPro: IPR000384; MGS.
CC DR Pfam: PF01808; AICARFT_IMPCHas; 1.
CC DR Pfam: PF02142; MGS; 1.
CC DR ProDom: PD004666; AICARFT_IMPCHas; 1.
CC DR Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme;
CC KW Complete proteome.
CC SEQUENCE 512 AA; 55739 MW; 7FD6B0FE5485234C CRC64;

```





```

AC P41885;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Hypothetical 111.7 kDa protein F37A4.7 in chromosome III.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton L., Waterston R.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 FYVE-TYPE ZINC FINGER.
CC -1- SIMILARITY: SOME, TO YEAST YNL087W.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
-----
CC CC
DR EMBL; U00032; AAA50635.1; -.
DR HSSP; P47709; 1ZBD.
DR WormPep; F37A4.7; CE00705.
DR InterPro; IPR000008; C2.
DR InterPro; IPR002149; LRI.
DR InterPro; IPR003315; RPH3A_effector.
DR InterPro; IPR001565; Synaptoagmin.
DR InterPro; IPR000306; znf_FYVE.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF02318; RPH3A_effector; 1.
DR PRINTS; PR00360; C2DOMAIN.
DR PRINTS; PR00399; SYNAPTOTAGMN.
DR SMART; SM00239; C2; 2.
DR PROSITE; PS00499; C2_DOMAIN_1; 2.
DR PROSITE; PS50004; C2_DOMAIN_2; 2.
DR PROSITE; PS50178; ZF_FYVE; 1.
KW Hypothetical protein; Repeat; Zinc-finger.
FT ZN_FING 168 233 FYVE-TYPE.
FT DOMAIN 758 847 C2 DOMAIN 1.
FT FT 898 987 C2 DOMAIN 2.
SO SEQUENCE 1021 AA; 111749 MW; 947C38B7C37BA462 CRC64;

Query Match          9.9%; Score 71; DB 1; Length 1021;
Best Local Similarity 28.6%; Pred. No. 71;
Matches 34; Conservative 16; Mismatches 39; Indels 30; Gaps 8;

QY 28 NYTGVDKSGNELLSPHYVEFPKRGTT---LTKEIEYV--EWALDATAYPEFRVEL 81
   |: :||| | | :||: ||| | | :||: ||| | | :||: ||| | | :||: ||| | |
Db 768 NLKAMDNSG----FSDPYVKFHLLPGNTKATKLTSKTIEKTLNPENMEMSY---GITED 821

QY 82 DPSAKI-EVTYYDKNK-----KK--EETKSFP|--TEKGFWVPDLSEHIKN 122
   | | | | | | :|: | | | | | | :|: | | | | | | :|: | | | | | | :|: |
Db 822 DKREKILRVTVLDRDRIGSDFLGETRIALLKLLNDNMKKFNLYESALPVPOQTKEEN 880

RESULT 41
TCPG_YEAST
ID TCGP_YEAST STANDARD; PRT; 534 AA.
AC P39077;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE T-complex protein 1, gamma subunit (TCP-1-gamma) (CCT-gamma).
DE CCT3 OR TCP3 OR BIN2 OR YUL014W OR J1336.
OS Saccharomyces cerevisiae (Baker's yeast).
```

```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94377502; PubMed-7916460;
RA Chen X., Sullivan D.S., Hufaker T.C.;
RT "Two yeast genes with similarity to TCP-1 are required for
RL microtubule and actin function in vivo.";
RN Proc. Natl. Acad. Sci. U.S.A. 91:9111-9115(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA To Van D., Perea J., Jacq C.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MOLECULAR CHAPERONE; ASSIST THE FOLDING OF
CC ATP HYDROLYSIS. KNOWN TO PLAY A ROLE, IN VITRO, IN THE FOLDING OF
CC ACTIN AND TUBULIN. IN YEAST MAY PLAY A ROLE IN MITOTIC SPINDLE
CC FORMATION.
CC -1- SUBUNIT: HETERO-OLIGOMERIC COMPLEX OF ABOUT 850 TO 900 kDa THAT
CC FORMS TWO STACKED RINGS, 12 TO 16 NM IN DIAMETER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; U09480; AAA21658.1; -.
CC EMBL; Z49289; CAA89305.1; -.
CC PIR; S48875; S48875.
CC HSSP; P48425; 1A6D.
CC SGD; S0003551; CCT3.
CC InterPro: IPR002423; TCP1_cpnl60.
CC InterPro: IPR002194; TCP_1.
CC Pfam; PF00118; cpnl60_TCP1_1.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00750; TCP1_1; 1.
CC PROSITE; PS00751; TCP1_2; 1.
CC PROSITE; PS00995; TCP1_3; 1.
CC Chapterone; ATP-binding; Multigene family.
CC FT CONFLICT 25 26 IT -> HA (IN REF. 1).
CC FT CONFLICT 86 89 DEEV -> ERGG (IN REF. 1).
CC FT CONFLICT 111 111 L -> F (IN REF. 1).
CC FT CONFLICT 265 265 K -> M (IN REF. 1).
CC FT CONFLICT 274 274 Q -> L (IN REF. 1).
CC FT CONFLICT 292 292 MISSING (IN REF. 1).
CC FT CONFLICT 478 478 G -> A (IN REF. 1).
CC SEQUENCE 534 AA; 58814 MW; 5C5AFF3D67D9A3B6 CRC64;
SQ

```



DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Thermostable DNA ligase (EC 6.5.1.1) (Polydeoxyribonucleotide synthase  
[ATP]) (lig(Tk)).  
GN lig.  
OS Pyrococcus kodakaraensis.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.  
OX NCBI\_TaxID=69014;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=KOD1;  
RX MEDLINE=20507812; PubMed=11053387;  
RA Nakatani M., Ezaki S., Atomi H., Imanaka T.;  
RT "A DNA ligase from a hyperthermophilic archaeon with unique cofactor  
specificity.";  
RL J. Bacteriol. 182:6424-6433(2000).  
CC -1- FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA  
RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA.  
CC -1- CATALYTIC ACTIVITY: ATP + (deoxyribonucleotide)(N) +  
(deoxyribonucleotide)(M) -> AMP + diphosphate +  
(deoxyribonucleotide)(N+M).  
CC -1- COFACTOR: REQUIRES MAGNESIUM. CAN ALSO ACTS WITH NAD AS A  
COFACTOR.  
CC -1- SUBUNIT: MONOMER.  
CC -1- MISCELLANEOUS: OPTIMUM PH IS 8.0. STILL ACTIVE AT 100 DEGREES  
CELSIUS.  
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AB042527; BAB15949.1; ALT\_INIT.  
DR InterPro: IPR000977; DNA\_ligase.  
DR Pfam: PF01068; DNA\_ligase; 1.  
DR PROSITE: PS00697; DNA\_LIGASE\_A1; FALSE\_NEG.  
DR PROSITE: PS00333; DNA\_LIGASE\_A2; FALSE\_NEG.  
DR PROSITE: PS50160; DNA\_LIGASE\_A3; 1.  
KW DNA repair; DNA replication; DNA recombination; Cell division; Ligase;  
KW ATP-binding; NAD.  
FT BINDING 249 249 AMP (BY SIMILARITY).  
SQ SEQUENCE 559 AA; 63748 MW; 91AB32542E03D20D CRC64;  
Query Match 9.9%; Score 70.5; DB 1; Length 559;  
Best Local Similarity 29.7%; Pred. No. 38;  
Matches 22; Conservative 13; Mismatches 30; Indels 9; Gaps 3;  
QY 42 PHYVEF-PI-KPGTTLTKKEIYVEMALDATAKKEFRVELDPSAKIEVTYYDKKKKE 99  
DB 456 PHSGEFLPVGKVGSGFTDEDLVEFTKMLKPIYVQEGKFEIEPKFVIEVTY-----Q 508  
QY 100 ETKSPITEKGFV 113  
DB 509 EIQSPKPKYKSGFAL 522  
RESULT 43  
YD56\_YEAST  
ID YD56\_YEAST STANDARD; PRT; 608 AA.  
AC 004399;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Putative multicopper oxidase YDR506C (EC 1.-.-.-).  
GN YDR506C OR D9719.12.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;

RN [1]  
RP SEQUENCE FROM N.A.  
RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E.,  
RA Berne A., Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,  
RA Hunnicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,  
RA Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,  
RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,  
RA Winant A., Yelton M., Botstein D., Davis R.W.;  
RL Submitted (Aug-1995) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: COULD BE A IRON TRANSPORT MULTICOPPER OXIDASE, WHICH IS  
REQUIRED FOR FERROUS IRON HIGH AFFINITY UPTAKE. MAY BE REQUIRED TO  
OXIDIZE FE(II) AND RELEASE IT FROM THE TRANSPORTER. ESSENTIAL  
COMPONENT OF COPPER-DEPENDENT IRON TRANSPORT (BY SIMILARITY).  
CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH  
CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE  
2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: TO IRON TRANSPORT MULTICOPPER OXIDASES.  
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.  
CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U33057; AAB64948.1; -.  
DR SGD: S0002914; YDR506C.  
DR InterPro: IPR001117; Cu-oxidase.  
DR InterPro: IPR002355; MulticCu\_oxidase2.  
DR Pfam: PF00394; Cu-oxidase; 3.  
DR PROSITE: PS00079; MULTICOPPER\_OXIDASE1; 2.  
DR PROSITE: PS00080; MULTICOPPER\_OXIDASE2; 1.  
KW Hypothetical protein; Transport; Oxidoreductase; Copper; Repeat;  
KW Metal-binding; Iron transport.  
FT DOMAIN 51 163 PLASTOCYANIN-LIKE 1.  
FT DOMAIN 243 374 PLASTOCYANIN-LIKE 2.  
FT DOMAIN 421 548 PLASTOCYANIN-LIKE 3.  
FT METAL 100 100 COPPER (TYPE 2) (BY SIMILARITY).  
FT METAL 102 102 COPPER (TYPE 3) (BY SIMILARITY).  
FT METAL 145 145 COPPER (TYPE 3) (BY SIMILARITY).  
FT METAL 147 147 COPPER (TYPE 3) (BY SIMILARITY).  
FT METAL 452 452 COPPER (TYPE 1) (BY SIMILARITY).  
FT METAL 455 455 COPPER (TYPE 2) (BY SIMILARITY).  
FT METAL 457 457 COPPER (TYPE 3) (BY SIMILARITY).  
FT METAL 530 530 COPPER (TYPE 3) (BY SIMILARITY).  
FT METAL 531 531 COPPER (TYPE 1) (BY SIMILARITY).  
FT METAL 532 532 COPPER (TYPE 3) (BY SIMILARITY).  
FT METAL 536 536 COPPER (TYPE 1) (BY SIMILARITY).  
FT METAL ? ? COPPER (TYPE 1) (BY SIMILARITY).  
SQ SEQUENCE 608 AA; 69281 MW; D51B830F15E704AC CRC64;  
Query Match 9.9%; Score 70.5; DB 1; Length 608;  
Best Local Similarity 29.0%; Pred. No. 42;  
Matches 29; Conservative 16; Mismatches 38; Indels 17; Gaps 6;  
QY 41 SPHYVEPIKPGTTLTKKEIYVEMALDATAKKEFRVELDPSA--KIEVYY--DKNK 96  
DB 355 NPNDVSTIEHLPGFT---KAELYRDIEPTQENKKLR-TKADPVAVFEFDYAYKDESTK 409  
QY 97 KKEETKSPITEKGFVVPDLSEHIKNP---GFNLITKVI 133  
DB 410 QKYGTMKVNERTF-----SEYVKDPVRFGEFNETYDIYI 444  
RESULT 44  
YD46\_MYCPN  
ID YD46\_MYCPN STANDARD; PRT; 115 AA.  
AC P75432;  
DT 16-OCT-2001 (Rel. 40, Created)

```

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MPN346 (H91_orf115).
OS MPN346 OR MP490.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000048; AAB96138.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 115 AA; 13468 MW; 95E6EBEF9F054B CRC64;

Query Match          9.8%; Score 70; DB 1; Length 115;
Best Local Similarity 25.3%; Pred. No. 6.5;
Matches 25; Conservative 13; Mismatches 31; Indels 30; Gaps 3;

QY 54 TLTRKEIEYYVEWALD-----ATAYKE-FRVVELDPSAKIEVYYDKKK 97
   ||::|::|::||      || | : | : | : ||
Db 29 TLMERIKHIIVDTLDKYLDATONTPKFDNFENPVKGFNSIFACESKEAIEYKFEKK 88
   : | | | | :       || | | | : ||

QY 98 KEETKSFPITEKGFWVPDLSEHTKNPGFNLTGVTEKK 136
   : | | | | :       || | | | : ||
Db 89 QTEGKKLPKIASIIAQ-----TKVMNKK 113
   : | | | | :       || | | | : ||

RESULT 45
Y255_METUA STANDARD; PRT; 251 AA.
ID Y255_METUA
AC Q57703;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0255.
GN MJ0255.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uutterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kalne B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC EMBL; U67480; AAB98242.1; .
DR TIGR; MJO255; .
DR InterPro; IPR003830; DUF210.
DR Pfam; PF02679; DUF210; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 251 AA; 28368 MW; 9B3F2CD349880E4C CRC64;

Query Match          9.8%; Score 70; DB 1; Length 251;
Best Local Similarity 25.5%; Pred. No. 16;
Matches 40; Conservative 19; Mismatches 44; Indels 54; Gaps 9;

QY 29 VTGVDSKGNELLSPHYVEFPIKP-----GTT-----LTKEIEYVENALDA 70
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15 LTVYLDKG--LPPKFVEDYLYKCGDYIDVFKFGWGTSAYIDRDVYKEKINYKDWGIKV 71

QY 71 -----TAYKE-----FRVEL-DPSAKIEVITYYDKNKKKEETKSFP- 106
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 72 YPGCTLFEYAYSKGKFDFLNCCEKLGFEAVEISDSSDSLSEERKNAIKRAKDNQFVVL 131

QY 107 TEKGFVVPD-----LSEHIKNPGFNL---ITKVVIE 134
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 132 TEVGKKMPDKDKQLTIDDRIKLINFDLDAGADYVITE 168

RESULT 46
KEMS_FSVMD
ID KEMS_FSVMD STANDARD; PRT; 978 AA.
AC P00345; Q86597;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein fms (EC 2.7.1.112).
GN V-FMS.
OS Feline sarcoma virus (strain McDonough).
OC Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
OC NCBI_TaxID=11778;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84119469; Pubmed=6582485;
RA Hampe A., Gobet M., Sherr C.J., Galibert F.;
RT "Nucleotide sequence of the feline retroviral oncogene v-fms shows
RT unexpected homology with oncogenes encoding tyrosine-specific protein
RT kinases.";
RT Proc. Natl. Acad. Sci. U.S.A. 81:85-89(1984).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RX MEDLINE=92015516; Pubmed=1833563;
RA Smola U., Hennig D., Hadwiger-Fangmeyer A., Schuetz B., Pfaff E.,
RA Niemann H., Tamura T.;
RT "Reassessment of the v-fms sequence: threonine phosphorylation of the
RT COOH-terminal domain.";
RL J. Virol. 65:6181-6187(1991).
CC -1- FUNCTION: V-FMS IS DERIVED FROM THE RECEPTOR FOR COLONY
CC STIMULATING FACTOR 1 (CSF-1).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-FMS
CC POLYPROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-
CC PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; K01643; AAA43045.1; -  
 DR EMBL; S59588; AAB20028.1; -  
 DR PIR; A00654; TVMVMD.  
 DR HSP; P11362; 1FGK.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003598; Ig\_C2.  
 DR InterPro; IPR003600; Ig\_Like.  
 DR InterPro; IPR001824; Receptor\_tyr\_kin\_III.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00047; Ig; 5.  
 DR Pfam; PF00069; pkinase; 2.  
 DR SMART; SM00410; Ig\_Like; 4.  
 DR SMART; SM00408; IGC2; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS00240; RECEPTOR\_TYR\_KIN\_III; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Polypeptide; Tyrosine-protein kinase; Oncogene; Transferase; Receptor;  
 KW Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;  
 KW Immunoglobulin domain; Repeat.  
 FT DOMAIN 1 543 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 544 568 POTENTIAL.  
 FT DOMAIN 569 978 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 58 138 IG-LIKE C2-TYPE DOMAIN 1.  
 FT DOMAIN 141 231 IG-LIKE C2-TYPE DOMAIN 2.  
 FT DOMAIN 238 332 IG-LIKE C2-TYPE DOMAIN 3.  
 FT DOMAIN 333 431 IG-LIKE C2-TYPE DOMAIN 4.  
 FT DOMAIN 432 536 IG-LIKE C2-TYPE DOMAIN 5.  
 FT DOMAIN 613 942 PROTEIN KINASE.  
 FT NP\_BIND 619 627 ATP (BY SIMILARITY).  
 FT BINDING 647 647 ATP (BY SIMILARITY).  
 FT DISULFID 76 118 POTENTIAL.  
 FT DISULFID 161 211 POTENTIAL.  
 FT DISULFID 258 312 POTENTIAL.  
 FT DISULFID 451 516 POTENTIAL.  
 FT MOD\_RES 841 841 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 973 973 PHOSPHORYLATION.  
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 524 524 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 714 714 L -> P (IN REF. 1).  
 FT CONFLICT 971 978 QRTPPVAR -> RGPPL (IN REF. 1).  
 SQ SEQUENCE 978 AA; 108491 MW; 4C7CAC4835185EBF CRC64;

Query Match 9.8%; Score 70; DB 1; Length 978;  
 Best Local Similarity 27.1%; Pred. No. 83;  
 Matches 26; Conservative 18; Mismatches 42; Indels 10; Gaps 5;

QY 19 EPTGP---YLAVNTGVDSKGNELLSPHYVEPIKPGTTLTKERK-EYVEMALDATAK 74  
 DB 33 EAMGPRLVLLMATAWHAQGVPIQSPGPELVPEGTVTTLRCVNGSVEM--DGPISP 90  
 QY 75 EFRVVELDPSAKIEVTYYDKNKKKEFSFPITEKG 110  
 DB 91 HMN-LDIDPPSSILTT---NNATFQNTGYHCTEPG 122

RESULT 47  
 NXP2\_MOUSE  
 ID NXP2\_MOUSE STANDARD; PRT; 244 AA.  
 AC 061199;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Neurexophilin 2 precursor (Fragment).  
 GN NXP2 OR NPH2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-96285495; PubMed-8699246;  
 RA Petrenko A.G., Ullrich B., Missler M., Krasnoperov V., Rosahl T.W.,  
 RA Suedhof T.C.;  
 RT "Structure and evolution of neurexophilin";  
 RL J. Neurosci. 16:4360-4369(1996).  
 CC -1- FUNCTION: MAY BE SIGNALING MOLECULES THAT RESEMBLE NEUROPEPTIDES  
 CC AND THAT ACT BY BINDING TO ALPHA-NEUREXINS AND POSSIBLY OTHER  
 CC RECEPTORS (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE NEUREXOPHILIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 9.7%; Score 69.5; DB 1; Length 244;  
 Best Local Similarity 24.7%; Pred. No. 18;  
 Matches 41; Conservative 22; Mismatches 42; Indels 61; Gaps 11;

QY 7 GRYKK---GDDASYFEP-----TGP-----YLMVNTGVDSKGNELLS-P-H 43  
 DB 86 GFRKMGWGDHSHNIKYKLNLLITGKIYDHGNGTFSVYFRHNSTGLGNVSLSVPPSK 145  
 QY 44 YVEFPIKPGTTL-TKEKIEY--YVEMALDATAVKEFRVVELDPSAKIEVYYDKNKKKEE 100  
 DB 146 YVEFEISPOSTLETESKSFNCHIEYEKTDRA-KKTALCNEDPS---KICY-----QEQ 195  
 QY 101 TKS-----FPITEKGFVVPDLSEHIKNP 123  
 DB 196 TQSHVSWLCSKPKFYICHIIFYSVDYKLVQK--VCPDYNHYSERP 239

RESULT 48  
 GYRB\_TREDE  
 ID GYRB\_TREDE STANDARD; PRT; 638 AA.  
 AC 087545;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)



```

DE  DNA gyrase subunit B (EC 5.99.1.3).
GN  gyrb.
OS  Treponema denticola.
OC  Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX  NCBI_TaxID=158;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 35405;
RA  Greene S.R., Stamm L.V.;
RT  "Characterization of the gyrb region of Treponema denticola.";
RL  Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC  STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC  INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC  DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC  -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC  of double-stranded DNA.
CC  -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC  BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
CC  ENZYME FORMS AN A2B2 TETRAMER.
CC  -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: AF083949; AAC62070.1; -.
DR  HSSP: P06982; 1AUG.
DR  InterPro: IPR002288; DNA_gyraseB_C.
DR  InterPro: IPR001241; DNA_topoisomI.
DR  InterPro: IPR003594; HATPase_C.
DR  InterPro: IPR002936; Toprim.
DR  Pfam: PF00986; DNA_gyraseB_C; 1.
DR  Pfam: PF00204; DNA_topoisomI; 1.
DR  Pfam: PF02518; HATPase_C; 1.
DR  Pfam: PF01751; Toprim; 1.
DR  PRINTS: PR00418; TP12FAMILY.
DR  ProDom: PD000616; DNA_topoisomI; 1.
DR  ProDom: PD149633; DNA_gyraseB_C; 1.
DR  SMART: SM00387; HATPase_C; 1.
DR  SMART: SM00433; TOP2c; 1.
DR  PROSITE: PS00177; TOPOISOMERASE_II; 1.
DR  Topoisomerase; Isomerase; ATP-binding.
KW  SEQUENCE 638 AA; 71391 MW; AD0DFFA091C2D5A6 CRC64;
SQ

```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (XTCAD-1).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95324376; PubMed=7600966;
RA Levine E., Lee C.H., Kintner C., Gumbliner B.M.;
RT "Selective disruption of E-cadherin function in early Xenopus embryos
RT by a dominant negative mutant.";
RL Development 120:901-909(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Tail bud;
RX MEDLINE=94368839; PubMed=8086449;
RA Tooi O., Fujii G., Tashiro K., Shiohara K.;
RT "Molecular cloning of cDNA for XTCAD-1, a novel Xenopus cadherin, and
RT its expression in adult tissues and embryos of Xenopus laevis.";
RL Blochm. Biophys. Acta 1219:121-128(1994).
RN [3]
RP SEQUENCE OF 149-872 FROM N.A.
RX MEDLINE=94363396; PubMed=8081882;
RA Broders F., Girault J.M., Simonneau L., Thlery J.P.;
RT "Sequence and distribution of Xenopus laevis E-cadherin transcripts.";
RL Cell Adhes. Commun. 1:265-277(1993).
RN [4]
RP SEQUENCE OF 149-169.
RX MEDLINE=91347911; PubMed=1879345;
RA Angres B., Mueller A.H.J., Kellermann J., Hausen P.;
RT "Differential expression of two cadherins in Xenopus laevis.";
RL Development 111:829-844(1991).
CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS: CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. E-CADHERIN IS A LIGAND FOR
CC INTEGRIN ALPHA-E/BETA-7.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: NON-NEURAL EPITHELIAL TISSUES.
CC -1- DEVELOPMENTAL STAGE: APPEARS IN THE EMBRYONIC ECTODERM DURING
CC GASTRULATION WHEN EPIDERMAL DIFFERENTIATION COMMENCES AND IT
CC DISAPPEARS FROM THE NEURAL PLATE AREA UPON NEURAL INDUCTION.
CC -1- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U04708; AAA93116.1; -
DR EMBL; L29057; AAA61489.1; -
DR EMBL; X75454; CAA53206.1; -
DR PIR; A60128; A60128.
DR HSSP; P09803; 1SUH.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR00233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS50268; CADHERIN_2; 4.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 148
FT CHAIN 149 872 EPIITHELIAL-CADHERIN.

```



```

FT DOMAIN 149 701 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 702 722 POTENTIAL.
FT DOMAIN 723 722 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 723 722 CADHERIN 1.
FT DOMAIN 244 254 CADHERIN 2.
FT DOMAIN 358 368 CADHERIN 3.
FT DOMAIN 577 587 POLY-LEU.
FT DOMAIN 710 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 242 242 I -> V (IN REF. 3).
FT CONFLICT 332 332 N -> T (IN REF. 3).
FT CONFLICT 487 487 E -> R (IN REF. 3).
FT CONFLICT 501 502 AT -> CS (IN REF. 3).
FT CONFLICT 539 541 GNG -> EMA (IN REF. 3).
FT CONFLICT 548 548 K -> R (IN REF. 3).
FT CONFLICT 557 557 V -> G (IN REF. 3).
FT CONFLICT 567 567 P -> L (IN REF. 3).
FT CONFLICT 602 604 GFR -> EPQ (IN REF. 3).
FT CONFLICT 638 640 GOS -> DK (IN REF. 2).
FT CONFLICT 640 647 SILELRPK -> VYLSDL (IN REF. 3).
FT CONFLICT 660 660 T -> A (IN REF. 3).
FT CONFLICT 770 770 A -> S (IN REF. 3).
FT CONFLICT 842 842 P -> S (IN REF. 2 AND 3).
FT CONFLICT 870 871 DE -> GED (IN REF. 3).
SQ SEQUENCE 872 AA; 96065 MW; 08637967EEFB4664 CRC64;

```

Query Match 9.7%; Score 69.5; DB 1; Length 872;  
 Best Local Similarity 25.5%; Pred. No. 81;  
 Matches 28; Conservative 17; Mismatches 46; Indels 19; Gaps 4;

```

OY 8 KYRKGDDASYE---PTGPY-LMNVNTGVDSKGNELLSPHYV-----EPIKPG 52
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 409 KIIRKGNENYFSIQDTGNIGLTKYKGLDYE---LKKQYILSVITNKANESVPLQTS 464
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 53 TLTTRKEIEYVWALDATAYKEFRVVELDPSAKIEVYTYDKNKKKEETK 102
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 465 TATVTVSVEDVNEAPIFLPVPKEVSVSEDLPSGVATYTTADDPDKEQNQ 514
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 50
ACON_LEGPN STANDARD; PRT; 891 AA.
AC P37032;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Aconitase hydratase (EC 4.2.1.3) (Citrate hydro-lyase) (Aconitase)
DE (Major iron-containing protein) (MICP) (IP210).
GN ACN.
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Legionellaceae; Legionella.
OX NCBI_TaxID=446;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PHILADELPHIA 1;
RX MEDLINE=93374864; PubMed=8366052;
RA Mengaud J.M., Horwitz M.A.;
RT "The major iron-containing protein of Legionella pneumophila is an
RT aconitase homologous with the human iron-responsive element-binding
RT protein."
RL J. Bacteriol. 175:5666-5676(1993).
CC -1- CATALYTIC ACTIVITY: citrate -> cis-aconitrate + H(2)O.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- MISCELLANEOUS: ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE
CC (3FE-4S) FORMS. THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE
CC CATALYTIC SITE THAT INTERCONVERTS CITRATE, CIS-ACONITASE, AND
CC ISOCITRATE.

```

```

CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L22081; AAA25295.1; -.
CC PIR: B48642; B48642.
DR HSSP: P16276; 7ACN.
DR InterPro: IPR001030; Aconitase.
DR InterPro: IPR000573; Aconitase_C.
DR Pfam: PF00330; aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PRINTS: PR00415; ACONITASE.
DR PRODOM: PD000511; Aconitase; 1.
DR PROSITE: PS00450; ACONITASE_1; 1.
DR PROSITE: PS01244; ACONITASE_2; 1.
KW Lyase; Tricarboxylic acid cycle; Iron-sulfur; 4Fe-4S.
FT METAL 435 435 IRON (IRON-SULFUR CLUSTER)
FT METAL 501 501 (BY SIMILARITY).
FT METAL 501 501 IRON (IRON-SULFUR CLUSTER)
FT METAL 504 504 (BY SIMILARITY).
FT METAL 504 504 IRON (IRON-SULFUR CLUSTER)
FT METAL 504 504 (BY SIMILARITY).
SQ SEQUENCE 891 AA; 98209 MW; F81FC99520859B48 CRC64;

```

Query Match 9.7%; Score 69.5; DB 1; Length 891;  
 Best Local Similarity 19.3%; Pred. No. 83;  
 Matches 41; Conservative 25; Mismatches 51; Indels 95; Gaps 8;

```

OY 7 GYRKGDDASYEPTGPIYLMVNT-----GVDSKGNELLSPHYVEFPKPGTTLT----- 56
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 258 GKLKEG-----ITAFDVLVLTVMRLRKGVGKFEVEFGPLNDPLADRATISNMAP 310
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 57 -----KEKIEYVWALDATAYKEFRVVELDPSAKIEVYTYDKN----- 95
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 311 EYGATCGFPPVDKETIKY-----LELTGRDKHTIALVEAYAKAOGMWDKDNPEVFTDS 365
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 96 -----KKKEETKSPITEKGF----- 111
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 366 LHUDLGSVEPSLAGPKRPQDKVNLSSLPVEFNNFLIEVGKEKEKEKETPAVKNKDFOMKHG 425
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 112 --VVPDLSE--HIKNGF---NLITKVIEK 135
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 426 HVIYAITSCTNTSNPSVLMAGLVAKKAIEK 457
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: September 1, 2002, 03:27:20  
 Job time: 511 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2002, 03:17:47 ; Search time 69.56 Seconds  
(without alignments)  
338.230 Million cell updates/sec

Title: US-09-728-670-10  
Perfect score: 715  
Sequence: 1 SSSFDKGGKYYKGDASYFEP.....SEHIKNPGFNLITKVIEKK 136

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	163	16 Q99SU7	Q99SU7 staphylococ
2	710	99.3	163	2 Q9LC46	Q9LC46 staphylococ
3	703	98.3	163	2 Q9AM04	Q9AM04 staphylococ
4	699	97.8	136	2 Q33929	Q33929 staphylococ
5	228	31.9	377	2 Q934F7	Q934F7 streptococ
6	138.5	19.4	416	2 Q9ZFE2	Q9ZFE2 streptococ
7	88	12.3	569	10 Q9ZRH6	Q9ZRH6 zea mays (m
8	87	12.2	348	16 Q34353	Q34353 bacillus su
9	84.5	11.8	483	9 Q9G004	Q9G004 bacterioph
10	83.5	11.7	606	16 Q9CMA9	Q9CMA9 pasteurilla
11	83	11.6	611	16 Q9X0Y4	Q9X0Y4 thermotoga
12	82.5	11.5	16215	5 Q9NFS3	Q9NFS3 drosophila
13	81	11.3	526	2 Q9ROS7	Q9ROS7 mycoplasma
14	80.5	11.3	158	17 Q59474	Q59474 pyrococcus
15	80	11.2	543	17 Q971U1	Q971U1 sulfolobus
16	79.5	11.1	378	5 Q95QB3	Q95QB3 caenorhabdi

17	78	10.9	350	10 Q39980	Q39980 hysocyanus
18	77.5	10.8	1586	10 Q9LPB1	Q9LPB1 arabidopsis
19	77	10.8	559	10 Q65323	Q65323 capsicum an
20	77	10.8	559	10 Q81923	Q81923 capsicum an
21	77	10.8	608	11 Q99J21	Q99J21 mus musculu
22	76.5	10.7	289	17 Q29463	Q29463 archaeoglob
23	76.5	10.7	349	17 Q971G3	Q971G3 sulfolobus
24	76	10.6	252	16 Q9CFA0	Q9CFA0 lactococcus
25	75	10.5	933	11 Q91X78	Q91X78 mus musculu
26	74.5	10.4	666	2 P96018	P96018 staphylococ
27	74	10.3	233	5 Q9XUQ2	Q9XUQ2 caenorhabdi
28	74	10.3	233	17 Q29007	Q29007 archaeoglob
29	74	10.3	256	5 Q9XUQ3	Q9XUQ3 caenorhabdi
30	74	10.3	462	10 Q04622	Q04622 arabidopsis
31	74	10.3	480	10 Q9M156	Q9M156 arabidopsis
32	74	10.3	552	16 Q9PPW0	Q9PPW0 ureaplasma
33	73.5	10.3	257	16 Q92N17	Q92N17 rhizobium m
34	73.5	10.3	431	10 Q9XIR2	Q9XIR2 arabidopsis
35	73.5	10.3	764	3 Q42976	Q42976 schizosacch
36	73.5	10.3	792	11 Q91X25	Q91X25 mus musculu
37	73.5	10.3	1536	16 Q9X079	Q9X079 thermotoga
38	73	10.2	171	11 Q9CUN9	Q9CUN9 mus musculu
39	73	10.2	345	13 Q91946	Q91946 gallus gall
40	73	10.2	419	17 Q973T1	Q973T1 sulfolobus
41	73	10.2	470	10 Q22732	Q22732 arabidopsis
42	73	10.2	654	10 Q9FRF0	Q9FRF0 oryza sativ
43	73	10.2	875	4 Q9H706	Q9H706 homo sapien
44	73	10.2	2505	11 Q63577	Q63577 rattus norv
45	73	10.2	2509	4 Q16702	Q16702 homo sapien

ALIGNMENTS

RESULT	ID	Q99SU7	PRELIMINARY;	PRT;	163 AA.
AC	Q99SU7	Q99SU7	Q99SU7	Q99SU7	Q99SU7
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	STAPHYLOKINASE PRECURSOR.				
GN	SAK OR SAI758.				
OS	Staphylococcus aureus (strain N315).				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Staphylococcus.				
OX	NCBI_TaxID=158879;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=21311952; PubMed=11418146;				
RA	Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,				
RA	Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,				
RA	Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,				
RA	Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,				
RA	Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,				
RA	Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,				
RA	Ogasawara N., Hayashi H., Hiramatsu K.;				
RT	"Whole genome sequencing of methicillin-resistant Staphylococcus				
RT	aureus."				
RL	Lancet 357:1225-1240(2001).				
DR	EMBL: AP003135; BABA3032.1; -.				
DR	HSSP: P00802; 1SSN.				
DR	InterPro: IPR004093; Staphylokinase.				
DR	Pfam: PF02821; Staphylokinase; 1.				
KW	Complete proteome.				
SQ	SEQUENCE 163 AA; 18520 MW; 413CDAE14BE9FD40 CRC64;				

Query Match 100.0%; Score 715; DB 16; Length 163;  
Best Local Similarity 100.0%; Pred. No. 3.4e-59;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SSSFDKGGKYYKGDASYFEP TGPYLNVNTGVDSKGNELLSPHYVEFP IKGTTLTKEKI 60

```
DB 28 SSSFDKGGKKGGDASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKI 87
QY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSFPITEKGFVVPDLSEHI 120
DB 88 EYVEMALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSFPITEKGFVVPDLSEHI 147
QY 121 KNPGENLITKVIEKK 136
DB 148 KNPGENLITKVIEKK 163

RESULT 2
Q9LC46 PRELIMINARY; PRT; 163 AA.
AC Q9LC46;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE STAPHYLOKINASE.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20219005; PubMed=10754251;
RA Horii T., Yokoyama K., Barua S., Odagiri T., Futamura N., Hasegawa T.,
RA Ohta M.;
RT "The staphylokinase gene is located in the structural gene encoding N-
RT acetylmuramyl-L-alanine amidase in methicillin-resistant
RT Staphylococcus aureus.";
RL EMBL; AB033232; BAA95011.1; -.
DR HSSP; P00802; ISSN.
DR InterPro; IPR004093; Staphylokinase.
DR Pfam; PF02821; Staphylokinase; 1.
KW kinase.
SQ SEQUENCE 163 AA; 18474 MW; 3DB45E35046029DD CRC64;

Query Match 99.3%; Score 710; DB 2; Length 163;
Best Local Similarity 98.5%; Pred. No. 9.9e-59;
Matches 134; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSSFDKGGKKGGDASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKI 60
DB 28 SSSFDKGGKKGGDASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKI 87
QY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSFPITEKGFVVPDLSEHI 120
DB 88 EYVEMALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSFPITEKGFVVPDLSEHI 147
QY 121 KNPGENLITKVIEKK 136
DB 148 KNPGENLITKVIEKK 163

RESULT 3
Q9AM04 PRELIMINARY; PRT; 163 AA.
AC Q9AM04;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE STAPHYLOKINASE SAKXH.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RA Wei W., Xiang H., Tan H.;
```

```
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF332619; AAK11636.1; -.
DR HSSP; P00802; ISSN.
KW Kinase.
SQ SEQUENCE 163 AA; 18509 MW; 0A266B5EEAF65B5A CRC64;

Query Match 98.3%; Score 703; DB 2; Length 163;
Best Local Similarity 98.5%; Pred. No. 4.4e-58;
Matches 134; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSSFDKGGKKGGDASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKI 60
DB 28 SSSFDKGGKKGGDASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKI 87
QY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSFPITEKGFVVPDLSEHI 120
DB 88 EYVEMALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSFPITEKGFVVPDLSEHI 147
QY 121 KNPGENLITKVIEKK 136
DB 148 KNPGENLITKVIEKK 163

RESULT 4
Q93929 PRELIMINARY; PRT; 136 AA.
AC Q93929;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE STAPHYLOKINASE (FRAGMENT).
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 29213;
RX MEDLINE=97417758; PubMed=9271816;
RA Kim S.H., Chun H.S., Han M.H., Park N.Y., Suk K.;
RT "A novel variant of staphylokinase gene from Staphylococcus aureus
RT ATCC 29213.";
RL Thromb. Res. 87:387-395(1997).
DR EMBL; U77328; AAB84174.1; -.
DR HSSP; P00802; ISSN.
DR InterPro; IPR004093; Staphylokinase.
DR Pfam; PF02821; Staphylokinase; 1.
KW kinase.
FT NON_TER
SQ SEQUENCE 136 AA; 15551 MW; 79916BB136CA1A3F CRC64;

Query Match 97.8%; Score 699; DB 2; Length 136;
Best Local Similarity 97.8%; Pred. No. 8.5e-58;
Matches 133; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SSSFDKGGKKGGDASYFEPTGPLYMNVNTGVDKGNELLSPHYVEFPPIKPGTTLTKEKI 60
DB 1 SSSFDKGGKKGGDASYFEPTGPLYMNVNTGVEGKENELLSPHYVEFPPIKPGTTLTKEKI 60
QY 61 EYVEMALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSFPITEKGFVVPDLSEHI 120
DB 61 EYVEMALDATAYKEFRVVELDPSAKIEVTTYDKNKKKEETKSFPITEKGFVVPDLSEHI 120
QY 121 KNPGENLITKVIEKK 136
DB 121 KNPGENLITKVIEKK 136

RESULT 5
Q934F7 PRELIMINARY; PRT; 377 AA.
ID Q934F7
```



AC Q93AF7; 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE PLASMINOGEN ACTIVATOR PRECURSOR.  
GN PAUB.  
OS Streptococcus uberis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1349;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SK880;  
RA Ward P.N., Leigh J.A.;  
RT "Characterisation of Paub, a novel broad spectrum plasminogen  
activator from Streptococcus uberis.";  
RL J. Bacteriol. 0:0-0(0).  
DR EMBL; AJ314852; CAC85651.1; -.  
KW Signal.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 377 PLASMINOGEN ACTIVATOR.  
SQ SEQUENCE 377 AA; 43841 MW; 9283C8AA97EF5E4F CRC64;

Query Match 31.9%; Score 228; DB 2; Length 377;  
Best Local Similarity 35.8%; Pred. No. 2.5e-13;  
Matches 44; Conservative 30; Mismatches 43; Indels 6; Gaps 1;

QY 14 DASFEPTGPLYMNVNTGVDSKGNELSPHYVEFPKPGTTLTKKEIEYVEMALDATA 73  
DB 30 EVNYQEPQGTMLTINITGSEKQGLLSPEYMQFRLTAGEVINKNELDKVQLVIDSAAS 89  
QY 74 KEFRVVELDPSAKIEVYDYDKNKKKEETKSPITEKGFVVDLSEHIKNPGFNLIKVI 133  
DB 90 NQFEVVDKPKESKVEWSYFDRHLLD-----ITERGFIVPDYSIYEKKKPSFLLTGPIVI 143  
QY 134 EKK 136  
DB 144 QRK 146

RESULT 6  
ID Q92FE2 PRELIMINARY; PRT; 416 AA.  
AC Q92FE2; 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE STREPTOKINASE (FRAGMENT).  
GN SKC.  
OS Streptococcus equisimilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=119602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=87-542-W;  
RX MEDLINE=20038313; PubMed=10569766;  
RA Caballero A.R., Lottenberg R., Johnston K.H.;  
RT "Cloning, expression, sequence analysis, and characterization of  
streptokinases secreted by porcine and equine isolates of  
streptococcus equisimilis.";  
RT streptococcus equisimilis.";  
RL Infect. Immun. 67:6478-6486(1999).  
DR EMBL; AF104301; AAC83933.1; -.  
DR InterPro; IPR004093; Staphylokinase.  
DR Pfam; PF02821; Staphylokinase; 1.  
KW Kinase.  
FT NON\_TER 416 416  
SQ SEQUENCE 416 AA; 47869 MW; CAB79D18496BCCC4 CRC64;

Query Match 19.4%; Score 138.5; DB 2; Length 416;  
Best Local Similarity 28.1%; Pred. No. 6.3e-05;

Matches 38; Conservative 27; Mismatches 49; Indels 21; Gaps 4;  
QY 7 GKYYKGGDA-----SYFEPTGPLYMNVNTGVDSKGNELSPHYVEFPKPGTTLTKKEIE 61  
DB 47 GTYQPTDPTFVNSKDYQDTGTGLYLMINVGVDNRHNTLVLPQYQDIKIDYKNTYTKDELL 106  
QY 62 YVVEWALDPAVAYKEFRVVELDPSAKIEVYDYD-----KNKKKEETKSPITEKG 110  
DB 107 RFVQNVIDA-SMENFKVIDLHEVKLVRTATQGE MSPNTFSLAENQYNDANOSYLLGNV 165  
QY 111 FVVP-----DLSEHIK 121  
DB 166 LVEPVSKRNLEKVIK 180

RESULT 7  
ID Q92R86 PRELIMINARY; PRT; 569 AA.  
AC Q92R86; 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE PDI-LIKE PROTEIN.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Panicoidae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99026300; PubMed=9808743;  
RA Laughner B.J., Sehne P.C., Ferl R.J.;  
RT "A novel nuclear member of the thioredoxin superfamily.";  
RL Plant Physiol. 118:987-996(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Laughner B., Ferl R.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U90944; AAD04231.1; -.  
DR InterPro; IPR004146; DC1.  
DR InterPro; IPR000063; ThioRed.  
DR Pfam; PF03107; DC1; 1.  
DR Pfam; PF00085; ThioRed; 2.  
DR PROSITE; PS00194; THIOREDOXIN; UNKNOWN\_1.  
SQ SEQUENCE 569 AA; 63356 MW; 5472A3991DCE3300 CRC64;

Query Match 12.3%; Score 88; DB 10; Length 569;  
Best Local Similarity 28.4%; Pred. No. 4.7;  
Matches 31; Conservative 19; Mismatches 39; Indels 20; Gaps 5;  
QY 23 PYLMVNTGVDSKGNELSPHYVEFPKPGTTLTKKEIEYVEMALDATAVK 74  
DB 97 PVLAVPFS--DSEGREALDGOFKVSGIFHLVILDAKTGEVYTEDGVEFVSEYGEVAYPFT 154  
QY 75 EFRVVELDPSAKIEVYDYDKNKKKE-----ETKSPITEKGFVVDLSE 118  
DB 155 PDRINELKEQKAE-----KENQTIQSVLVTSTRDYLLISNKGDKVP-ISE 198

RESULT 8  
ID Q34353 PRELIMINARY; PRT; 348 AA.  
AC Q34353; 01-JAN-1998 (Tremblrel. 05, Created)  
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
DE YDJN PROTEIN.  
GN YDJN.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MARBURG 168;
RA Kasahara Y., Nakai S., Ogasawara N., Yata K., Sadale Y.;
RL DNA Res. 0:0-0(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
  Azevedo V., Bertero M.G., Besiieres P., Bolotin A., Borchert S.,
  Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
  Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
  Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
  Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
  Entlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
  Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
  Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
  Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
  Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
  Jorls B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
  Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
  Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
  Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
  Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
  Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
  Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
  Presacan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
  Rieger M., Rivoita C., Rocha E., Roche B., Rose M., Sadale Y.,
  Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
  Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
  Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
  Takeuchi M., Yamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
  Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
  Viari A., Wambult R., Wedler E., Wedler H., Weltzenegger T.,
  Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
  Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
  RT "The complete genome sequence of the gram-positive bacterium Bacillus
  subtilis."
  RL Nature 390:249-256(1997).
  [3]
  RP SEQUENCE FROM N.A.
  RC STRAIN-168;
  RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
  RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
  DR EMBL; AB007638; BAA22769.1; -.
  DR EMBL; Z99107; CAB12445.1; -.
  KW Complete proteome.
  SQ SEQUENCE 348 AA; 39710 MW; F297D69D4C72967D CRC64;

Query Match          12.2%; Score 87; DB 16; Length 348;
Best Local Similarity 26.5%; Pred. No. 3.2;
Matches 31; Conservative 18; Mismatches 34; Indels 34; Gaps 5;

QY 5 DKGKYYKGGDASYEPTGYLMVNVTVGDSKGNELLSPHYVEFPITKPGTTLTKKEIEYV 64
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 27 DKG-HEKADVSNTESGDELVSITDID-----LITKYY-----ENDKVIHEKLTSTP 75

QY 65 EWALDATAIKERVELDPSAKIEVTYTDKNKKKEETKSFPITEKGFVVPDLSEHIK 121
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 76 AFALD-----OKQOVLYTGNNEQNEMLFKL-----DLKSHK 109

RESULT 9
Q9G004 PRELIMINARY; PRT; 483 AA.
AC Q9G004;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ORF41.
OS bacteriophage phi ETA.
```

```

OC Viruses.
OX NCBI_TaxID=106284;
RN [1]
RP SEQUENCE FROM N.A.
RA Sugai M., Yamaguchi T., Hayashi T., Nakasone K., Takami H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20566787; PubMed=11115106;
RA Yamaguchi T., Hayashi T., Takami H., Nakasone K., Ohnishi M.,
  Nakayama K., Yamada S., Komatsuzawa H., Sugai M.;
  RT "Phage conversion of exfoliative toxin A production in Staphylococcus
  aureus."
  RL Mol. Microbiol. 38:694-705(2000).
  DR EMBL; AP001553; BAA97627.1; -.
  SQ SEQUENCE 483 AA; 55712 MW; DDBB2547E0DCEC80 CRC64;

Query Match          11.8%; Score 84.5; DB 9; Length 483;
Best Local Similarity 25.5%; Pred. No. 8.2;
Matches 27; Conservative 21; Mismatches 27; Indels 31; Gaps 5;

QY 29 VTGVDKGNELLSPHYVE-----FPIKPG----TTLTKKEIEYVEWALDATAIKEF 76
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 139 LTGASNKGIEWLHPYLDDEEGEFKLFVRPAEGGIPITWDKEHEELFAFI----- 186

QY 77 RVELDPSAKIEVTYTDKNKKKEETKSFPITEKGFVVPDLSEHIK 122
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 RMYKLENETKVE--YMDK-----VTVNVYVYENGSLIPDYSNNLEN 225

RESULT 10
Q9CMA9 PRELIMINARY; PRT; 606 AA.
AC Q9CMA9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FADD.
GN FADD_2 OR PM0925.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
  RT "Complete genomic sequence of Pasteurella multocida pm70."
  RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
  DR EMBL; AE006132; AAK03009.1; -.
  DR InterPro; IPR000873; AMP-bind.
  DR PRINTS; PR00154; AMPBINDING.
  DR PROSITE; PS00455; AMP_BINDING; 1.
  KW Complete proteome.
  SQ SEQUENCE 606 AA; 69594 MW; FB500E49401E5224 CRC64;

Query Match          11.7%; Score 83.5; DB 16; Length 606;
Best Local Similarity 23.3%; Pred. No. 13;
Matches 38; Conservative 28; Mismatches 62; Indels 35; Gaps 8;

QY 1 SSSFDKGYKGGDASYEPTGP-YLMVNVTVGDSKGN-ELLSPHYVEFPITKPGTTLTKE 58
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 434 ADSFTEDGFLKTGDAGEFDPGNLXITDRIKELMKTSGKYIAPOYIE-----TKVGKD 487

QY 59 KIEYVWALDATAIKERFV-----ELDPSAK-IEVTYYDK-----NK 96
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 488 KFIQIAYIADAKKYVSALIVPCFNSLEFYAKQILNIKYHDRLELIKHSIDILQMEQRI 547

QY 97 KKEETKSFPITEKGFVVPDL-----SEHIKNPGFNLTQVIEK 135
   |||::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

DB 548 LQKELPSFEQIKKFTLLPOAFTTKMEI-TPTLKRKVILLER 589

RESULT 11

Q9X0Y4 PRELIMINARY; PRT; 611 AA.

AC Q9X0Y4; 01-NOV-1999 (Tremblrel. 12, Created)

DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE CONSERVED HYPOTHETICAL PROTEIN.

GN TM1257.

OS Thermotoga maritima.

OC Bacteria; Thermotogales; Thermotoga.

OX NCBI\_TaxID=2336;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MSB8 / DSM 3109;

RX MEDLINE=99287316; PubMed=10360571;

RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;

RA \*Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.\*

RT Nature 399:323-329(1999).

RL EMBL; AJ271740; CAB93524.1; -.

DR TIGR; TM1257; -.

KW Complete proteome.

SO SEQUENCE 611 AA; 70010 MW; 7FB042D4E49097AB CRC64;

Query Match 11.6%; Score 83; DB 16; Length 611;

Best Local Similarity 21.5%; Pred. No. 15;

Matches 28; Conservative 27; Mismatches 45; Indels 30; Gaps 5;

QY 8 KYKKGDDASYFEPTG-----PYLMVNTGVDSKGN-----ELLSPHYVEPIKP 51

DB 163 RYILGDNGAHINVSQSGVAAKTSYTTFLVKSMIETSSKNDGDLRELREARYITFNKVG 222

QY 52 GTTLTKKIEIYVEMALDATAKKE-FRVVELDPSAKIEIYTYDKNKKKEETKSFPITEKG 110

DB 223 ESLLELDRIIS--KEWYSEKREKWDENRVLGIEPKPFENAVFY-----APSRKEKG 269

QY 111 FVVPDLSEHI 120

DB 270 AYIPDVNKRRL 279

RESULT 12

Q9NFS3 PRELIMINARY; PRT; 16215 AA.

AC Q9NFS3; 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE D-TITIN.

GN SLS OR D-TITIN OR CG1915.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Zhang Y.Q., Broadie K.S.;

RT \*Characterization of Drosophila D-Titin gene.\*

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ271740; CAB93524.1; -.

DR HSSP; P56276; ITLK.

DR FlyBase; FBgn0003432; sls.

DR InterPro; IPR002106; AA\_trna\_ligase\_II.

DR InterPro; IPR003961; FN\_III.

DR InterPro; IPR003598; Ig\_c2.

DR InterPro; IPR003600; Ig\_1like.

DR InterPro; IPR003006; Ig\_MHC.

DR InterPro; IPR001452; SH3.

DR Pfam; PF00041; fn3; 5.

DR Pfam; PF00047; Ig; 48.

DR Pfam; PF00018; SH3; 1.

DR SMART; SM00408; IGC2; 15.

DR SMART; SM00410; IG\_1like; 34.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS00179; AA\_TRNA\_LIGASE\_II\_1; UNKNOWN\_1.

DR PROSITE; PS50002; SH3; 1.

KW Immunoglobulin domain.

SO SEQUENCE 16215 AA; 1841509 MW; 242C8765E00F7603 CRC64;

Query Match 11.5%; Score 82.5; DB 5; Length 16215;

Best Local Similarity 25.5%; Pred. No. 8.2e+02;

Matches 36; Conservative 32; Mismatches 52; Indels 21; Gaps 8;

QY 8 KYKKGDDASYFE-----PTGPYLMVNTGVDSKG-NELLSPHYVEPIKPGTTLTK 57

DB 10444 RHEKGEVQVLEIVESVAPGEEPFYEINVISSANTEGDESEITTDKIK--KSRKIKK 10500

QY 58 EKIEYVEMALDATAKKEFRVELDPSAKIEIYTYDKNKKKEETKS-FPITEKGFVVP-D 115

DB 10501 DDLDAYIQQLINA-----EIPVTELEKYEKIDYDGAKAKPKKLKAKTKPKPIIDEGETLQVG 10556

QY 116 LSEHIKNPGFNLTFRVIEKK 136

DB 10557 VTEH--EPTKKLTKRKEEK 10575

RESULT 13

Q9RQST PRELIMINARY; PRT; 526 AA.

AC Q9RQST; 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE MAJOR LIPOPROTEIN.

GN LPPA.

OS Mycoplasma mycoides subsp. mycoides LC.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscites;

OC Entomoplasmataceae.

OX NCBI\_TaxID=44100;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-Y-GOAT;

RX MEDLINE=99167669; PubMed=1006658;

RA Monnerat M.P., Thiaucourt F., Poveda J.B., Nicolet J., Frey J.;

RT \*Genetic and serological analysis of lipoprotein Lppa in Mycoplasma mycoides subsp. mycoides LC and Mycoplasma mycoides subsp. capri.;"

RT Clin. Diagn. Lab. Immunol. 6:224-230(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-Y-GOAT;

RX MEDLINE=99440592; PubMed=10512041;

RA Monnerat M.P., Thiaucourt F., Nicolet J., Frey J.;

RT \*Comparative analysis of the lppa locus in Mycoplasma capricolum subsp. capricolum and Mycoplasma capricolum subsp. capripneumoniae.;"

RL Vet. Microbiol. 69:157-172(1999).

DR EMBL; AF072714; AAF06069.1; -.

KW Lipoprotein.

SO SEQUENCE 526 AA; 60288 MW; 605CED85B652249D CRC64;

Query Match 11.3%; Score 81; DB 2; Length 526;

Best Local Similarity 23.7%; Pred. No. 19;

Matches 41; Conservative 24; Mismatches 64; Indels 44; Gaps 6;

QY 2 SSFDKGYKKGD-----DAS-----YFEPGDPYL-----MVNVTG 31  
: | | | | | : | | | | | : | | | | |  
Db 218 TEIDKNKKNKKDYIKPKETVDSRLSLGYSILAYMLLYAENTNNYKSLQETDKDAINFEG 277  
QY 32 VDSKGNELSPHYVEFPKPGTTLTKETIEYVEMALDATAK----EFVVELDPSAKI 87  
: | | | | | : | | | | | : | | | | |  
Db 278 LINKPTNLFNDKVFESVG-----TKELLFDENENYKRLYVKLVGAGFDINDINGTLTKV 332  
QY 88 EVTYVDKNKKKEETKSPITEKGFPVDPDSEHIKNPGENLIT-----KVIEK 135  
: | | | | | : | | | | | : | | | | |  
Db 333 EINNSENKEKEPISKEFSFGFRKVNTPDEPSKNPFYVSLTPADLKKITDK 385  
RESULT 14  
O59474 PRELIMINARY; PRT; 158 AA.  
ID O59474  
AC O59474:  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE HYPOTHETICAL 18.1 KDA PROTEIN PH1810.  
GN PH1810.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayasi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
DR EMBL; AP000007; BAA30929.1; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 158 AA; 18113 MW; 761428F02518AF06 CRC64;  
Query Match 11.3%; Score 80.5; DB 17; Length 158;  
Best Local Similarity 27.0%; Pred. No. 5.1;  
Matches 33; Conservative 16; Mismatches 38; Indels 35; Gaps 5;  
QY 38 ELSPHYVEFPKPGTTLTKK-----IEYVEMALDATAKKEFRVELDPS 84  
: | | | | | : | | | | | : | | | | |  
Db 4 DILKPEYIAFLIKEREKLLKKEKAQTIEIPGSEYEIEATVEYVDSTFERFIYIARSEDS 63  
QY 85 AKIEVTYVDKNKKKEETKSPITEKGFPVDP-----LSEHIKN-----PGFNLTIKV 131  
: | | | | | : | | | | | : | | | | |  
Db 64 ILVA-----RSSEKINTKG----EKKFLVKDKKGLKRLLLSEISKSKVGTGINFILAT 114  
QY 132 VI 133  
: |  
Db 115 II 116  
RESULT 15  
O971U1 PRELIMINARY; PRT; 543 AA.  
ID O971U1  
AC O971U1:  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HYPOTHETICAL PROTEIN ST1287.  
GN ST1287.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.  
OX NCBI\_TaxID=111955;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 10545 / 7;  
RX PubMed=11572479;  
RA Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermophilic  
RT Crenarchaeon, Sulfolobus tokodaii strain 7.";  
RL DNA Res. 8:123-140(2001).  
DR EMBL; AP000985; BAB66329.1; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 543 AA; 62412 MW; 1614BE43E6806CB7 CRC64;  
Query Match 11.2%; Score 80; DB 17; Length 543;  
Best Local Similarity 25.6%; Pred. No. 25;  
Matches 31; Conservative 21; Mismatches 55; Indels 14; Gaps 5;  
QY 6 KGKYYKGDASYFEPTGYLMVNTGVDKGNELSPHY-VEFPKPGTTLTKETIEYV 64  
: | | | | | : | | | | | : | | | | |  
Db 45 RNKVKLDDILSILSELG----IELSETDLREMAKYLPEYDVEFLKDKGLIKPHV---- 96  
QY 65 EWALDATA-YKEFRVELDPSAKIEVT---YYDKNKKKEETKSPITEKGFPVDPDSEHI 120  
: | | | | | : | | | | | : | | | | |  
Db 97 -FLDIIDKYKDKGILKDKOSKVVYTPDYYYQIKRRLLEENGLKIKDGLDVKELNINF 155  
QY 121 K 121  
: |  
Db 156 K 156  
RESULT 16  
O95QB3 PRELIMINARY; PRT; 378 AA.  
ID O95QB3  
AC O95QB3:  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HYPOTHETICAL 43.3 KDA PROTEIN.  
GN T12C9.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabdillidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Pauley A.;  
RT "The sequence of C. elegans cosmid T12C9.";  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U41104; AK18975.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 378 AA; 43299 MW; 0D11BF594E86C0B6 CRC64;  
Query Match 11.1%; Score 79.5; DB 5; Length 378;









```
DB 519 SFDEQFQTLQCVTASDNGDPPLSSNSLTFLVLDONDNTPEILYPSLPTDGSTGVELA 578
OY 118 EHIKNPGFNLTIRVV 132
DB 579 PRSAEPGY-LVTRVV 592

RESULT 26
ID P96018 PRELIMINARY; PRT; 666 AA.
AC P96018;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MECA PROTEIN.
GN MECA.
OS Staphylococcus sciuri.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1296;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC29062;
RX MEDLINE=97302569; PubMed=9158816;
RA Wu S., Piscitelli C., de Lencastre H., Tomasz A.;
RT "Tracking the evolutionary origin of the methicillin resistance gene:
RT cloning and sequencing of a homologue of meca from a methicillin-
RT susceptible strain of Staphylococcus sciuri.";
RL Microb. Drug Resist. 2:435-441(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC29062;
RA Wu S.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y09223; CAA70425.1; -.
DR InterPro; IPR001460; Transpeptidase.
DR Pfam; PF00905; Transpeptidase; 1.
SQ SEQUENCE 666 AA; 75606 MW; CF046F9D10F29529 CRC64;

Query Match 10.4%; Score 74.5; DB 2; Length 666;
Best Local Similarity 22.9%; Pred. No. 1e+02;
Matches 33; Conservative 23; Mismatches 53; Indels 35; Gaps 6;

OY 27 VNVGTGVDKSGNELSPHYVEFP-----IKPGTTLTK-----EKI---EYV----- 64
DB 141 INIEPLKSERGIIDRNVELATGTGTHEVGIVPNVSTSDYKAIKALDLSESYIKQQT 200
OY 65 --EWALDATAVKEFRVELDPSAKIEVYDYDKNKKKEETKSFPITER-----GFVVPDLIS 117
DB 201 EQDWVKDDTFVPLKTVQDMNODLKNFVEKHLTSQETESROYPLEEATTHLLGYGPINS 260
OY 118 EHIKNPGFN-----LITKVIEK 135
DB 261 EELKQKAFKGYKKDAIVGKKGIK 284

RESULT 27
ID P9XUQ2 PRELIMINARY; PRT; 233 AA.
AC P9XUQ2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE T08G5.9 PROTEIN.
GN T08G5.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RA Smye R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81589; CAB04722.1; -.
SQ SEQUENCE 233 AA; 25912 MW; F40BCED0A2A1710E CRC64;
```

```
Query Match 10.3%; Score 74; DB 5; Length 233;
Best Local Similarity 20.7%; Pred. No. 33;
Matches 25; Conservative 28; Mismatches 50; Indels 18; Gaps 4;
```

```
OY 20 PTGPYLMVNTGVDKSGNELSPHYVEFPPIKPGTTLTKKIEYVEMALDATAKEFRV 79
DB 91 PEPALRSNVTSLTSESDDIQGFLE-----NESL-FQIEYKIDYKNYRDYTI 139
OY 80 ----ELDPSAKIEVYDYDKNKKKEETKSFPITERGFVVPDLSEHIKNPGFNLTIRVIEK 135
DB 140 PSEGTVIPKSKCEIALICRPTGKSESDQMTISYK---TVDSNDNLANGCFSEMEKIVIDI 196
OY 136 K 136
DB 197 K 197
```

```
RESULT 28
ID O29007 PRELIMINARY; PRT; 233 AA.
AC O29007;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MEDIDUM-CHAIN ACYL-COA LIGASE (ALKK-4).
GN AF1261.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artlach P., Raine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE001017; AAB89982.1; -.
DR TIGR; AF1261; -.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 2.
KW Hypothetical protein; Ligase; Complete proteome.
SQ SEQUENCE 233 AA; 27066 MW; 14C1ECAR62093FB6 CRC64;
```

```
Query Match 10.3%; Score 74; DB 17; Length 233;
Best Local Similarity 22.7%; Pred. No. 33;
Matches 32; Conservative 24; Mismatches 41; Indels 44; Gaps 7;
```



```

QY 2 SSFDDKKYKKGGDDASYFEETGPLYMV--NVTGVDKSGKNELLS----- 41
      :||| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 97 NSFVDGWMKSG-DVAVITPEG-YVKVDRVKDVIKSGGEMISSVDLENLGMHPAVLEAC 154

QY 42 -----PHYVEFPI-----KPGTTLTKKEI-----EYVEWALDATAYKEFRVELDPSA 85
      | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 155 VVAAEHPKQWERPIAIVPKPGSEVTKDELREFLAKRFKQOLPDD-----IIFVNEIP 208

QY 86 KIEVTTYDDKNKKKEETKSFPI 106
      | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 209 KTSVGKFDKKRLRREQYRRYLI 229

RESULT 29
Q9XUQ3 PRELIMINARY; PRT; 256 AA.
ID Q9XUQ3;
AC Q9XUQ3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE T08G5.8 PROTEIN.
GN T08G5.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Smye R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT Investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z81589; CAB04721.1; -.
SQ SEQUENCE 256 AA; 28494 MW; 6BDFDFA05921D939 CRC64;

Query Match 10.3%; Score 74; DB 5; Length 256;
Best Local Similarity 20.7%; Pred. No. 37;
Matches 25; Conservative 28; Mismatches 50; Indels 18; Gaps 4;

QY 20 PTGPYLMVNTGVDSKGNELSPHYVEFPKPGTTLTKKEIYYEWALDATAYKEFRV 79
      | :| | | :| | :| | :| | :| | :| | :| | :| | :| |
Db 114 PEPAFLRSNVTSLTSESDDIQFFLE-----NESL-FQIEYKIDYKNRYDTIV 162

QY 80 ----ELDPSAKIEVTTYDDKNKKKEETKSPITEKGFVVPDLSEHIKNPGFNLTQVIEK 135
      :| :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 163 PSFGTVIPKSKCEIALICRPTGKSESDDQMTISYK--TVSDNDNLANGCFSEMEKIVIDI 219

QY 136 K 136
      |
Db 220 K 220

RESULT 30
ID 004622 PRELIMINARY; PRT; 462 AA.
AC 004622;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE A_IG002N01.15 PROTEIN.
GN A_IG002N01.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid1; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]

```

```

RP  SEQUENCE FROM N.A.
RC  STRAIN-CV. COLUMBIA;
RA  Scheet P., Maggi L.;
RT  "The sequence of A. thaliana IG002N01.";
RL  Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN-CV. COLUMBIA;
RA  wash-U;
RT  "The A. thaliana Genome Sequencing Project.";
RL  Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN-CV. COLUMBIA;
RA  Waterston R.;
RL  Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF007269; AAB61023.1; .
DR  InterPro: IPR002213; UDPGT.
DR  Pfam; PF00201; UDPGT; 1.
DR  PROSITE; PS00375; UDPGT; 1.
SQ  SEQUENCE 462 AA; 50779 MW; F2D6BA0E70BF2DE7 CRC64;

Query Match 10.3%; Score 74; DB 10; Length 462;
Best Local Similarity 28.6%; Pred. No. 74;
Matches 38; Conservative 16; Mismatches 51; Indels 28; Gaps 7;

QY 6 KGYKKGGDDASYEPTGPLYMNVNTGVDSKNE-----LSPHYVEFPPIKPGT 53
   | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 206 KALQEPGLDKPPVYPVGP--LVNIGKQEAQKTESECECLKWDNQPLGSVLYVSF--GSGG 261
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 54 TLTKKIEIYVEMALD-ATAYKFRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFV 112
   ||| |:: | | | | : : | | | | | | | | | | | | | | | |
DB 262 TLTCQOLN--ELALGLADSEORFLMWIRSPSGIANSSYFDSHSQTD-----PLT---FL 310
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 VPDLSEHIKNPGF 125
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 311 PPGFLERTKKRGF 323

RESULT 31
O9M156 ID O9M156 PRELIMINARY; PRT; 480 AA.
AC O9M156;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PUTATIVE FLAVONOL GLUCOSYLTRANSFERASE.
ATAG01070.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBL_TaxID=3702;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RA Iamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RA "Full Length cDNA of gene AT4g01070 (GI:7267604).";
RT

```



OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,  
RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,  
RA Lucos S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,  
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).  
DR EMBL: AC006193; AAD38257.1; -.  
DR HSSP: P00799; ZAST.  
DR MEROPS: A01.UPW; -.  
DR InterPro: IPR002328; ADH\_zinc.  
DR InterPro: IPR001969; ASP\_protease.  
DR InterPro: IPR001461; Pepsin.  
DR Pfam: PF00026; asp; 1.  
DR PRINTS: PR00792; PEPsin.  
DR PROSITE: PS00059; ADH\_ZINC; 1.  
DR PROSITE: PS00141; ASP\_PROTEASE; UNKNOWN\_2.  
KW Oxidoreductase; zinc.  
SO SEQUENCE 431 AA; 46340 MW; 644FF17BA386ECA CRC64;

Query Match	10.3%	Score 73.5;	DB 10;	Length 431;
Best Local Similarity	26.3%	Pred. No. 76;		
Matches	35;	Conservative	17;	Mismatches 52;
				Indels 29;
				Gaps 6;

```

QY      11 KGDASYFEFTGPILMVNTGVDSKGNELLSPHY----VEFP1KPGTLLTKEKIEYVEW 66
        | | | | : : : | | | : | : | | : | |
Db      275 KRDPATYY----FLNLEAISVGSKKIQFTSTIFGTGEGNIVIDSGTLLTLLPSNFYYEL 329
        | | | | : : : | | | : | : | | : | |
QY      67 -ALDATAYKEFRVELDPSAKIEVTTYDYDNKKKEETKSPITEKGFVDPDLSEHIKNDGF 125
        :: | | : | | : : | | : : | | | | : | |
Db      330 ESWVASTIKAERV--QDPDGLISLCYRDS-----SFKVPDITVHEKGDV 373
        | | | | : : : | | | : | : | | : | |
QY      126 ---NLITKVIEK 135
        | | | | : : :
Db      374 KLGNLNTFVAAYSE 386

```

RESULT	35			
042976				
ID	042976	PRELIMINARY;	PRT;	764 AA.
AC	042976;			
DT	01-JAN-1999	(TREMBLrel. 09, Created)		
DT	01-JAN-1999	(TREMBLrel. 09, last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, last annotation update)		
DE	HYPOTHETICAL 85.4 KDA PROTEIN C20F10.07 IN CHROMOSOME II.			
OS	SPBC20F10.07.			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=972;			
RA	Lyne M., McDougall R., Rajandream M.A., Barrell B.G., Beck A.,			
RA	Reinhardt R.;			
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: SOME, TO YEAST YHR080C.			
DR	EMBL: AL021747; CAAL6847.1; -.			
DR	InterPro: IPR004182; GRAM.			
DR	pfam: PF02893; GRAM; 1.			
KW	Hypothetical protein.			
SEQ	SEQUENCE 764 AA; 85354 MW; 080990E5C7FA7B2 CRC64;			

Query Match	10.38; Score 73.5; DB 3; Length 764;
-------------	--------------------------------------

Best Local Similarity 23.0%; Pred. No. 1.5e+02;  
Matches 34; Conservative 16; Mismatches 49; Indels 49; Gaps 6;

RESULT	36		
Q91XZ5			
ID	Q91XZ5	PRELIMINARY;	PRT; 792 AA.
AC	Q91XZ5;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, last sequence update)		
DE	01-DEC-2001 (TREMBLrel. 19, last annotation update)		
DE	PROTOCADHERIN BETA 5.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BRAIN;		
RX	MEDLINE=99308636; PubMed=10380929;		
RA	Wu Q., Maniatis T.;		
RT	"A striking organization of a large family of human neural cadherin-		
RL	like cell adhesion genes.";		
RL	Cell 97:779-790(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BRAIN;		
RX	MEDLINE=20202599; PubMed=10716726;		
RA	Wu Q., Maniatis T.;		
RT	"Large exons encoding multiple ectodomains are a characteristic		
RT	feature of protocadherin genes.";		
RT	Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BRAIN;		
RX	MEDLINE=21154914; PubMed=11230163;		
RA	Wu Q., Zhang T., Cheng J.-F., Kim Y., Grimwood J., Schmutz J.,		
RA	Dickson M., Noonan J.P., Zhang M.Q., Myers R.M., Maniatis T.;		
RT	"Comparative dna sequence analysis of mouse and human protocadherin		
RT	gene clusters.";		
RL	Genome Res. 11:389-404(2001).		
DR	EMBL; AY013787; AAK26076.1; -		
SQ	SEQUENCE 792 AA; 87272 MW; F7A313143227BB74 CRC64;		

Query Matchn	10.3%;	Score 73.5;	DB 11;	Length 792;
Best Local Similarity	25.3%;	Pred. No. 1.6e+02;		
Matches 39;	Conservative 17;	Mismatches 55;	Indels 43;	Gaps 8
QY 13	DDASYFEETGPYLM-----	VNVTGVDSDKGN-----	ELLSPHYVEFPPIKPGT	53
	:	: :	:	
Db 445	DNAPAFSQTS-YTMFVRENNSPALHIGTISATSDSGSNAHITYSLRPHDQIALHSLI			503
QY 54	TLTKEKIEYVEMALDATAYKEFRV-----	ELDPSAKIEVYYDKNKKKE-----		99
	: : : :         :	:		
Db 504	SINADNGQLFALSALDYEALQGFEFFYVGATDRGSPELSSQALVRVVVLDDNDNAPFVLYP			563
QY 100	-ETKSEPTTEKGFVVPDLSEHIKNPGFNLTQVY			132
	:             :   :	:		
Db 564	LQNASAPCTE---VLPRAAE---PGY-LVTQVY			589

```
RESULT 37
O9X079 ID O9X079 PRELIMINARY; PRT; 1536 AA.
AC O9X079;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
GN HYPOTHETICAL 176.0 KDA PROTEIN.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Liner K.D., Garrett M.M.,
RA Stewart A.M., Sutton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima.";
DR EMBL; AE001760; AAD36063.1; -.
DR TIGR; TM0984; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1536 AA; 175955 MW; 71D493908F0CC1F4 CRC64;

Query Match
Best Local Similarity 10.3%; Score 73.5; DB 16; Length 1536;
Matches 36; Conservative 13; Mismatches 44; Indels 57; Gaps 6;

OY 3 SFDKGYKKGGDDASYEPTGYLMVNTGVDSKGNELSPHY-----VEFPKIP 51
Db 234 SFEGLYKA-----FEKT-----KTVSIFDTKKNEVYRSEFETDELGGFGSMKLPDTA 283
OY 52 GTTTLTK-----EKIEYVWALDATAKFRVVEL-----DP 83
Db 284 SVGLYKYNVDHGRRYEFVLEEVYRKPEYKVEITDKDVIYISGEVYNYLVKVFYNGQP 343
OY 84 SAKIEVTVYDKNKKKKEETKSPITEKGFV 113
Db 344 VAKAQVAVY-----VRAFPEEGSGYLV 365

RESULT 38
O9CUN9 ID O9CUN9 PRELIMINARY; PRT; 171 AA.
AC O9CUN9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE 4930430K04RIK PROTEIN (FRAGMENT).
GN 4930430K04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawal J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
```

```
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015254; BAB29766.1; -.
DR MGD; MGI:1921888; 4930430K04RIK.
DR InterPro; IPR000313; PWWP.
DR Pfam; PF00855; PWWP; 1.
DR SMART; SM00293; PWWP; 1.
DR NON_TER 1
SQ SEQUENCE 171 AA; 19652 MW; 4E9F0BB5BF60B7A6 CRC64;

Query Match
Best Local Similarity 10.2%; Score 73; DB 11; Length 171;
Matches 35; Conservative 15; Mismatches 38; Indels 52; Gaps 7;

OY 6 KGKYKKGGDDASYEPTGYLMVNTGVDSKGNELSPHYVEFPKIPGTT----- 54
Db 53 KGKY-----VT-YDQDN--VEKYVEFLGDPHSTAWMSAAFGHF 90
OY 55 -LTKEKIEYVWALDATAK-----FRVELDPSAKIEVTVYDK-NKKKETKS 103
Db 91 SLTLE-----AADCTKKRWYRSALAEAYOLYRCSAEQRLEVCCLSKPNRAKTDTKA 142
OY 104 FPITEKGFVVPDLSEHIKNP 123
Db 143 AVVTKKGMQVSKINTEKKRP 162

RESULT 39
O9I946 ID O9I946 PRELIMINARY; PRT; 345 AA.
AC O9I946;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SPINAL CORD-DERIVED GROWTH FACTOR.
GN SCDGF.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN; TISSUE=SPINAL CORD;
RX MEDLINE=20317014; PubMed=10858496;
RA Hamada T., U-Tel K., Miyata Y.;
RT "A novel gene derived from developing spinal cords, SCDGF, is a unique
RT member of the PDGF/VEGF family.";
RL FEBS Lett. 475:97-102(2000).
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; AB033829; BAB03265.1; -.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR00072; PDGF.
DR Pfam; PF00431; CUB; 1.
DR SMART; SM00042; CUB; 1.
DR SMART; SM00141; PDGF; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS50278; PDGF_2; 1.
SQ SEQUENCE 345 AA; 38940 MW; 97ACEA992BF5128C CRC64;
```





```
Db 477 LDAYVA---SEVKRQFKRPSPEKKIPIDPSVRNFFRGMASASVKEAIKLSYERTLKK-- 530
Oy 101 TKSEPITEKGFVVPDISE 118
Db 531 ----AYSCKSKTVPOLGE 544

RESULT 43
O9H706 PRELIMINARY; PRT; 875 AA.
AC 09H706;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CDNA: FLJ21610 FIS, CLONE COL07333.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK025263; BAB15094.1; .
DR InterPro: IPR001660; SAM.
DR SMART: SM00454; SAM; 1.
SQ SEQUENCE 875 AA; 97099 MW; 10A0E9147A9BFCB1 CRC64;

Query Match 10.2%; Score 73; DB 4; Length 875;
Best Local Similarity 30.4%; Pred. NO. 2e+02;
Matches 24; Conservative 15; Mismatches 20; Indels 20; Gaps 4;

Oy 43 HVEFPIKPGTTLTKKIEVYEWALDATAYKFRVVELDPSAKIEVYYDKNKKKEETK 102
Db 73 HVIIGP-----KIEIPVHYA-----GQFKLLEQDRDIKEPVOYF--NSVEEVAK 114

Oy 103 SFPITEKGFVVPDLSEHIK 121
Db 115 AFP--ERVYVMEITFNVK 131

RESULT 44
O63577 PRELIMINARY; PRT; 2505 AA.
AC 063577;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE FATY ACID SYNTHASE.
GN FATY ACID SYNTHASE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=89240686; PubMed=2717611;
RA Amy C.M., Witkowski A., Naggert J., Williams B., Randhawa Z.I.,
RA Smith S.;
RT "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT fatty acid synthase."
RL Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=92141210; PubMed=1736293;
```

```
RA Smith S., Naggert J.K., Williams-Ahlf B., Amy C.M.;
RT "Intron-exon organization of the gene for the multifunctional animal
RT fatty acid synthase."
RL Proc. Natl. Acad. Sci. U.S.A. 89:1105-1108(1992).
DR EMBL: M84761; AAA41145.1; .
DR InterPro: IPR001227; Acyltransf_domain.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR003880; Phosphopant_attach.
DR InterPro: IPR000051; SAM_bind.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00698; Acyl_transf; 1.
DR Pfam: PF00107; adh_zinc; 1.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF02801; ketoacyl-synt_C; 1.
DR Pfam: PF00550; pp-binding; 1.
DR Pfam: PF00975; Thioesterase; 1.
DR PROSITE: PS50075; ACP_DOMAIN; 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 2505 AA; 272663 MW; 4F97CFA5E83EB78 CRC64;
```

```
Query Match 10.2%; Score 73; DB 11; Length 2505;
Best Local Similarity 30.0%; Pred. NO. 6.9e+02;
Matches 24; Conservative 14; Mismatches 32; Indels 10; Gaps 4;

Oy 27 VNVTVGDSKGNELSPHYVEFPIKPGTTLTKKIEVYEWALDATAYKFRVVELDP--S 84
Db 803 VHLTGIDINPNALFPP--VEFVPRGTPL----ISPHIKW--DHSQTWDIPVAEDFPNGS 854

Oy 85 AKIEVYYDKNKKKEETKSF 104
Db 855 SSSSATYVYNIDASSESDHY 874

RESULT 45
O16702 PRELIMINARY; PRT; 2509 AA.
AC 016702;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE FATY ACID SYNTHASE (EC 2.3.1.85).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hennigar R.A., Jenner K.H., Helne H.S., Kayler A.E., Wood F.D.,
RA Kuhnajda F.P., Pasternack G.R.;
RT "Molecular Cloning of Tumor-Associated Human Fatty Acid Synthase."
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U29344; AAA73576.1; .
DR InterPro: IPR001227; Acyltransf_domain.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR003880; Phosphopant_attach.
DR InterPro: IPR000051; SAM_bind.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00698; Acyl_transf; 1.
DR Pfam: PF00107; adh_zinc; 1.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF02801; ketoacyl-synt_C; 1.
DR Pfam: PF00550; pp-binding; 1.
DR Pfam: PF00975; Thioesterase; 1.
DR PROSITE: PS50075; ACP_DOMAIN; 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.
KW Acyltransferase; Phosphopantetheine; Transferase.
SQ SEQUENCE 2509 AA; 273089 MW; 689018A8219C13B7 CRC64;
```







Tue Sep 3 09:15:52 2002

us-09-728-670-10.rspt

Page 19

```

QY      64 VEWALDATAVKEFRVVELDPSAKIEVTYYDKNKKKKEETKSP-IT EK-GFVVPD 115
      |      |  ||      |  ||      |  ||      |  ||      |  ||
Db 232 V-----KGGRVV-----YRANPYYYEKD---ENG NRLLPYIKEKGVILPD 268

```

Search completed: September 1, 2002, 03:27:09  
Job time: 562 sec





seq\_name: gb\_ba:AP003135  
seq\_documentation\_block: 291150 bp DNA 1linear BCT 07-FEB-2002  
LOCUS AP003135  
DEFINITION Staphylococcus aureus subsp. aureus N315 genomic DNA, complete  
genome, section 7/10.  
ACCESSION AP003135 BA000018  
VERSION AP003135.2 GI:14349227  
KEYWORDS  
SOURCE Staphylococcus aureus subsp. aureus N315 (sub\_species:aureus N315,  
strain:N315) DNA.  
ORGANISM Staphylococcus aureus subsp. aureus N315  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Staphylococcus.  
REFERENCE  
AUTHORS 1 (sites)  
Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,  
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,  
Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A.,  
Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C.,  
Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J.,  
Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,  
Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.  
TITLE Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus  
JOURNAL Lancet 357 (9264), 1225-1240 (2001)  
MEDLINE 21311952  
REFERENCE 2 (bases 1 to 291150)  
AUTHORS Aoki,K., Oguchi,A., Hosoyama,A., Nagai,Y., Kuroda,M., Hiramatsu,K.  
and Kikuchi,H.  
TITLE Direct Submission  
JOURNAL Submitted (30-JAN-2001) Kenichi Aoki, National Institute of  
Technology and Evaluation, Biotechnology Center, 2Chome 49-10  
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
(E-mail:aoki-kenichi2@meti.go.jp, URL:http://www.bio.nite.go.jp/  
Tel:81-3-3481-1972, Fax:81-3-3481-8424)  
COMMENT On Jun 12, 2001 this sequence version replaced gi:13701545.  
FEATURES  
source 1. 291150  
/organism="Staphylococcus aureus subsp. aureus N315"  
/strain="N315"  
/sub\_species="aureus N315"  
/db\_xref="taxon:158879"  
complement(357. .1205)  
/gene="SA1571"  
complement(357. .1205)  
/gene="SA1571"  
/note="ORFID:SA1571"  
/codon\_start=1  
/transl\_table=11  
/product="D-alanine aminotransferase"  
/protein\_id="BAB42839.1"  
/db\_xref="GI:13701546"  
/translation="MEKIFLNGEYVSPSEAKVSYNDRGYVFGDCIYERYRVNKGLET  
VTEHYERLRSANEIGLDLNSYEELIELSRKLVDMNQIETGAIYIQATRGVAERNHS  
FPTPEVERPAIVATKSYDRPYDHLNGVNGTVEDIRMLRCDIKSLNLGNVLAKEXA  
VKYNAVEAIOHGETVTEGSSSNAYAIKDGVIYTHPINNYILNGITRIYIKKIAEDYN  
IPFKEETFTVDFLKNADDEVIVSSISAEVTPVIKLDGEPIINDGKVGPIITROLQEGPEKY  
IESHSI"  
complement(1209. .2618)  
/gene="SA1572"  
complement(1209. .2618)  
/gene="SA1572"  
/note="ORFID:SA1572  
Xaa-His dipeptidase homolog"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB42840.1"  
/db\_xref="GI:13701547"  
/translation="MMKEKVGQYEDQIINDLKGLLATESVRDDAKASEDAPVGPGRK  
ALDYMETIAHRDGFTHDVIDHAGRIEAGKNDVLGILCHVDVVPAGDGDNSNPPEY  
VTEDAIARGTLDDKGPITIAIYAIKILEDMNVDWKRRIMIGITDEESDWKCTDRYF  
KTEEMPTLGFAPDAAEFPCIHGEKGITTFDLVQNKLTEDQDEPDYELITFKSGERYNMV

PDHAEARVLKENMTDVIQDFEYFLEONHLOGDSTVDSGLVLTVEGKAVHGMPSIG  
VNAGLYLLKFLASLNDNNAQAFVAFNSRYLEFNSDFGEKMKFHTDVGDTTNGV  
ITYDNENAGLFGINLRYPEGFEEKANDRFANEIQDYGFEVKGKGVQPPHYVDKNDPF  
VOKLVTAYRNQNDMTPEPTYIGGTYARNLDKGVAFGAFMSDEDLMHOKNEYITTKQ  
LFNATSIYLEAIYSLCEE"  
complement(3176. .3598)  
/gene="SA1573"  
complement(3176. .3598)  
/gene="SA1573"  
/note="ORFID:SA1573"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB42841.1"  
/db\_xref="GI:13701548"  
/translation="MAKNLFKAILGIGAVAAVLVTRKDSRDLKAENKYYKQDPOS  
YKDNAKDKATOLGTIANETIKEVKTNPKEYANRLKNPKAFPEEKSKEFTEDNKTDE  
SIEGKFDEGGAAPNNNLRIVTEEDLKKKNALSDE"  
complement(3615. .4310)  
/gene="SA1574"  
complement(3615. .4310)  
/gene="SA1574"  
/note="ORFID:SA1574  
hypothetical protein, similar to 16S pseudouridyate  
synthase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB42842.1"  
/db\_xref="GI:13701549"  
/translation="MRIDFLANMGVTRNEVKQLLKGVLNVNVEQVKSPTHIEPE  
NDKITVRGELIEYIENVYIMLNKPKGYISATEDHHSKTVIDLPEYOHLPVGRDL  
KDTGELLITNDGDFNHELMSPNKHVSKKEYISANPTEDDQAFKEGVTLDGKVK  
PALITYIDNQTSHTVITYEGKYHQVKRMFHSIQNEVLHLRRIKIADLELDSNDSGEYR  
LITENDFDKLNK"  
complement(4307. .5968)  
/gene="SA1575"  
complement(4307. .5968)  
/gene="SA1575"  
/note="ORFID:SA1575  
spore cortex protein homolog"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB42843.1"  
/db\_xref="GI:13701550"  
/translation="MSESKEMVRGFTLITISILITKVLGVLPFPENYLLIGQENMAP  
FTYAYAPYNAIAVATAGVPLAASKYVAKYNAIGAKVSOKFKSSFVMSITGVLG  
LVLYFLAPYISELTARNIHDKNGWSVDITWIRIISMVVITPVLATWRCIFQGYK  
SMGPTAVSEVTEQIARVIFILIGSYLVNVEDGSILLANGIAPAAVAGAILGIFTLW  
YWRKRKHNIIDRVESDITDIDVSYGKYEIIAYSIPFVYSNPLFNLVDQFTHN  
GALSLVGIPSQLDIFENMLNMSTNKIWMIPTSLSAGFAVSLPIYITKTFAGRLHEM  
HQIRTSIGVLMFTVPASIGIMALQPLETFVFGYDPIVLGDPNHDGRLLFYAP  
VALISLSVTFASMLKQAKFKFSYSWIHFAPKIFLYSFIIMLGVEIVFPLANLFLEPTKLG  
YLIIILGVTVGILITGTTIKTRLADEFLGEIPEKLRVRVRLR"  
6375. .7643  
/gene="SA1576"  
6375. .7643  
/gene="SA1576"  
/note="ORFID:SA1576"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAB42844.1"  
/db\_xref="GI:13701551"  
/translation="MYQTIITGGGPGSLMAAFAVASEQSSVLLIEKKKGLGRKLKISG  
GGRCNVTNRLPYAEIIRKINIPGNGKFLYSPSIEDNESIIDFESRGVLLKEEDHGRMF  
PVSNKAQDVVDLVTITIEROHVTIKEEAVSRIEVNTDQTFYHTQNNSYESHSLVIA  
TGTSVPTQGSTGDKYKFAQDLGHTITTELEPTEVPITSAEPFIKSNRLKGLSKOVEL  
SVLKNGKKRISHQMDLFTHFGISGPAALRCSQFVYKEQKQNDQOHSMAIDAFPEL  
NHEQLKQHTISLSDTPDKIINKSLHGLIEERYLLFMLQAGIDENNTSHLSNQOLN  
DLVNMFGFVFKVNGTLPIDKAFVTGGVSLKEIQPTMMSKIVPGLFLCGEVLDIHG  
YTGYNITSALVTVGHVAGLYAGHYSHTSME"  
complement(7760. .14320)

gene



```
/gene="SA1577"
complement(7760. .14320)
/gene="SA1577"
/note="ORFID:SA1577
hypothetical protein, similar to FmbB protein"
/codon_start=1
/transl_table=11
/protein_id="BAB42845.1"
/db_xref="GI:13701552"
/translation="MNLKKNKYSIKYKVGIFSTLIGTVLLSNPNGAQLTTDNV
OSDTNQATPVNSODTNVANNRGLANSQNTPNOSATNOSTNOLVHNHNGSIANQAT
PTSVQSTPSAQNNNHDTGNTTATETYSNANKKDVSNNTTLNPNKTNENGSGGHLT
LKEIQEDVRRSSDKPELVAIAEQASNRPKRSRAPADPNAPADPAAAGNGCAP
VAITAPYPTDTPNANNAGQAPNEVLSFDNGIRPSTNRSVSVTVVDNLPGFTLIN
GGKGVESHMVRTSMFDSGDAKNYQAQGNVIALGRKGNDRHGDENGIEKSLTVN
PNSLIEFEFTMTTKNYQVTLIKNADNDTVIAEKSVAYGPIWLFKVPENVSLK
IQFVKNDATDARGIYQLRDGYKYDFVDSIGLHSGSHVYVERTEPTATNNKEFT
VTTSLKNGNGEASFNTDDFVYKVOEGEVYVNSLTKEPSSNGVDMDNFTYD
AANRVITIKSTGGSGSNPARLMPDKILDKYKLRVNNVPTPRVTENDLTLYKTYQ
DFINSPAESHTVSTNPYTIIDIMNKDALQAEVDRIQADYTFASLIDFENDLKRRAOT
ILDENRNVPNKRVSQADIDSITNQMHLLIRSVDAENAVKKVDQMEDLVNQDEL
TDEKQAAIOVIEHKNEIIGNIGDQTTDGVTRIKDQIGTIGSDTATPVKPNNAK
AIRDKATKQREIINATPDATEDIDALNQLATDEDAIDVNTNATTNADVEIAKNNG
INTIGAVNPQVTHKQARDAINOATATKROQINSNREATQEEKNALNELTQATNHAL
EQINQATNADVDNAKGDLNAINPAPYTVVQOARDASDAQOIAHIAEINANPDAT
QEEQOAIIDRYNAAVTANTNITLANNADVEQVKTNAIQGIAITPATKVTADAKA
IDKSAETQHNTEFNNDATLEEQAQOQLDQAVATAKONINADTNOEVAQAKDOGT
ONIVVIOPAQVKTDAKNAVNDKAREAITNINATPGATREKQELINRVNTLKNRALY
DIGVSTTAMVNSIRDDAVNOIGAVQVTKQOTATGVLNLAATKQOINQONTNAT
BEKQVALNQDELATAIINNINQADIVNAVDQAQQLGTAIKAINAIQPNIVKKPAPALAI
NOHYNAKLAEINATPDATNDEKNAINTLNODROQAIESIKANTNAEVDQAATVAEN
NIDAVQVDVVKQOARDKITAEVAKRIEAVKQTPNATDEEKQAVNQINQLKQOAINQ
INQNTNDQVDTTNQAVNAIDNVEAEVYIKPKAIADIEKAVKEQOQIDNSLSDTN
EKEVASQALAKEKEKALAAIDQOATNSQVNOQATNGVSAIKIIOETKVPKPAAREKIN
OKANELRAKINQDKEATAEERQVALDKINEFVNQAMTDITNNRTNOQVDDTTSQALDS
IALVAPEHIYRAAADAVAKQOYEAKQOIEQAEHATDEEKQVALNQLANNEKIALQNI
NOAVTNNDVKRYETNGIATLKGVQPHIVYIKPEAQOAIKATAENQVESIKDTPHATVE
LDEANQLISDTLKQAQOEIENTNODAAVTDVRNQTIKAIQOIKRYARRKRALDSIEE
NNKNQDLAIRNTLDTQDERDVAIDTLKIVNTIKNDIAQNTNAEVDRTETDGNDI
KVLPRQVVKPAROSQVGVKAEQNALIDQSDLSEERLAKHLEQALNQAIQOIN
HADKTAQVNDQSIDAQNIISKIKPATVKTATALQIOINATNKLILKANNEATDEEQ
NIAIAQVEKELIKAKQOIASAVTNADVAYLLHDEKNEIREIEPYINRKASAREQLTTL
FNDKQAIENAIQATVEERNISILAIQNIYDTAIGIODRNSAQVDTASLNLQTIH
DLDVHPIKKPDAAKTINDDLARVATLVQNYRKVSNNRKADALKAITALKQMDLELKT
ARTNADVDVAVKRFENVALSDIEAVITEKENSILRIDNIAOQVAKFAIATPEOLAKY
KVLIDQYVADGNRMIDEDATLNDIKQHTQFIVDEILAIKLPAAEKVSPKEIQAPAKV
CTPIKKEETHESRKEVELPNTGSEGMJLPKEFALITGAALLARRRRTKNEKES"
complement(14646. .14957)
/gene="SA1578"
complement(14646. .14957)
/gene="SA1578"
/note="ORFID:SA1578"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42846.1"
/db_xref="GI:13701553"
/translation="MKSITTDLKNKLFESKPVQIVDVRTDEETANGYIPNAKLIPMD
```

## alignment\_scores:

Quality:	715.00	Length:	136
Ratio:	5.257	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

## alignment\_block:

US-09-728-670-10 x AP003135/rev ..

Align seg 1/1 to reverse of: AP003135 from: 1 to: 291150

```
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
208296 TCAAGTTCATTCGACAAAGAAATATATAAAAAAGCGATGACGCGAGCTTA 208247
```

```
17 rphegluProthrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
208246 TTTTGAACCAACAGGCCCGTATTGTGATGTAATGTGACTGAGTTGATA 208197
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
|||||
208196 GTAAAGGAAATGATGCTATCCCTCATGTATGTCGAGTTCTTATAAA 208147
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
|||||
208146 CCTGGACTACACTTACAAAGAAAAAATGCAATACTATGTCGATGGCC 208097
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
208096 ATTAGATGCGACACATATTAAGAGCTTAGAGTAGTTGAATTAGATCCA 208047
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
208046 GCGCAAGATCGAAGTCACCTTATTATGATGAAGATAAGAAAAAGAGAA 207997
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
|||||
207996 ACGAAGCTTTCCTTAAACAGAAAAAGTTTGTGTCCACAGATTATC 207947
117 rgluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
|||||
207946 AGACCATATTAACCCCTGATTCACCTTAATTACAAAGGTTGTTATAG 207897
134 IuLysLys 136
|||||
207896 AAAAGAAA 207889
seq_name: gb_pat:A17530
```

## seq\_documentation\_block:

```
LOCUS A17530 414 bp DNA linear PAT 12-APR-1994
DEFINITION Bacteriophage phi-C recombinant DNA for staphylokinase C-phi-C.
ACCESSION A17530
VERSION A17530.1 GI:512331
KEYWORDS staphylokinase.
SOURCE Bacteriophage phi-C.
ORGANISM Bacteriophage phi-C.
VIRUSES; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
REFERENCE 1 (bases 1 to 414)
AUTHORS .
TITLE EXPRESSION SIGNAL-PEPTIDE-FREE STAPHYLOKINASES
JOURNAL Patent: WO 9313209-A 2 08-JUL-1993;
FEATURES
source
1. 414
/organism="Bacteriophage phi-C"
/specific_host="Staphylococcus aureus"
/db_xref="taxon:35242"
1. 414
/codon_start=1
/transl_table=11
/product="staphylokinase SAK-C-phi-C"
/protein_id="CAA01336.1"
/db_xref="GI:512332"
/translation="MSSSEFDKGYKKGGDASYFEPPTGYLMVNTGVGDKGNELSPH
YVEEPIKPGTTLTKKEIEYVEMALDATAYKEFRVELDPSAKIEVTYYIDKNNKKEET
KSEFITEKGFVVPDLSEHIKNPGENLITKVIVIEKK"
```

## CDS

```
/codon_start=1
/transl_table=11
/product="staphylokinase SAK-C-phi-C"
/protein_id="CAA01336.1"
/db_xref="GI:512332"
/translation="MSSSEFDKGYKKGGDASYFEPPTGYLMVNTGVGDKGNELSPH
YVEEPIKPGTTLTKKEIEYVEMALDATAYKEFRVELDPSAKIEVTYYIDKNNKKEET
KSEFITEKGFVVPDLSEHIKNPGENLITKVIVIEKK"
4. 411
mat_peptide
/product="staphylokinase SAK-C-phi-C"
BASE COUNT 161 a 59 c 83 g 111 t
ORIGIN
```

## alignment\_scores:

Quality:	711.00	Length:	136
Ratio:	5.228	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	99.265

## alignment\_block:

US-09-728-670-10 x A17530 ..

Align seg 1/1 to: A17530 from: 1 to: 414

```
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
   |||||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAGGCGATGACGCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
   |||||||
54 TTTTGAACCAACAGCCCGCTATTGATGTAATGTGACTGAGTTGATG 103
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
   ::|||||
104 GTAAAGGAATGAATGCTATCCCTCATATGTCGAGTTTCCATTAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
   |||||||
154 CCTGGGACTACACTTACAAAAGAAAAATGAACTATGTGCAATGGCC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
   |||||||
204 ATTAGATCGCAGACAGCATATAAGAGATTAGAGTTGAATTAGATCCA 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
   |||||||
254 GCGCAAGATCGAAGTCACTTATTATGATAGAATAAGAAAAAGAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
   |||||||
304 ACGAAGTCTTCCCTATTAACAGAAAAAGTTTGTGTCCAGATTATATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
   |||||||
354 AGAGCATATTAATAAACCTGATTCACCTTAATTACAAAGGTTGTTAT 403
134 LuLysLys 136
   |||||||
404 AAAAGAAA 411
```

seq\_name: gb\_pat:AR037170

seq\_documentation\_block:

LOCUS AR037170 414 bp DNA linear PAT 29-SEP-1999

DEFINITION Sequence 3 from patent US 5801037.

ACCESSION AR037170

VERSION AR037170.1 GI:5955026

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 414)

AUTHORS Behnke,D., Schloft,B., Albrecht,S., Guhrs,K.-H. and Hartmann,M.

TITLE Expression of signal-peptide-free staphylokinases

JOURNAL Patent: US 5801037-A 3 01-SEP-1998;

FEATURES

BASE COUNT 161 a 59 c 83 g 111 t

ORIGIN

alignment\_scores:

Quality: 711.00

Ratio: 5.228

Percent Similarity: 100.000

Percent Identity: 99.265

alignment\_block:

US-09-728-670-10 x AR037170 ..

Align seg 1/1 to: AR037170 from: 1 to: 414

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17

```
|||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAGGCGATGACGCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
   |||||||
54 TTTTGAACCAACAGCCCGCTATTGATGTAATGTGACTGAGTTGATG 103
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
   ::|||||
104 GTAAAGGAATGAATGCTATCCCTCATATGTCGAGTTTCCATTAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
   |||||||
154 CCTGGGACTACACTTACAAAAGAAAAATGAACTATGTGCAATGGCC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
   |||||||
204 ATTAGATCGCAGACAGCATATAAGAGATTAGAGTTGAATTAGATCCA 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
   |||||||
254 GCGCAAGATCGAAGTCACTTATTATGATAGAATAAGAAAAAGAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
   |||||||
304 ACGAAGTCTTCCCTATTAACAGAAAAAGTTTGTGTCCAGATTATATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
   |||||||
354 AGAGCATATTAATAAACCTGATTCACCTTAATTACAAAGGTTGTTAT 403
134 LuLysLys 136
   |||||||
404 AAAAGAAA 411
```

seq\_name: gb\_pat:E02873

seq\_documentation\_block:

LOCUS E02873 804 bp DNA linear PAT 29-SEP-1997

DEFINITION DNA encoding staphylokinase(SAK).

ACCESSION E02873

VERSION E02873.1 GI:2171098

KEYWORDS JP 1991098595-A/2.

SOURCE Staphylococcus aureus.

ORGANISM Staphylococcus aureus

REFERENCE 1 (bases 1 to 804)

AUTHORS Hanada,K., Aoyama,J., Ikeda,Y., Yoshimoto,M., Yoshimura,H.,

TAMAKI,S. and Kondo,I.

TITLE PRODUCTION OF PEPTIDES IN GENE ENGINEERING

JOURNAL Patent: JP 1991098595-A 2 24-APR-1991;

COMMENT

OS Staphylococcus aureus

PN JP 1991098595-A/2

PD 24-APR-1991

PF 11-SEP-1989 JP 1989234874

PI HANADA KAZUNORI, AOYAMA JUNKO, IKEDA YOKO, YOSHIMOTO MAKOTO,

YOSHIMURA HIROMITSU, TAMAKI SHIGEO, KONDO ISAMU PC

C12P21/00,C12N1/21,C12N15/62,C12N15/70,C12N15/75,(C12P21/00, PC

C12R1:19),

PC (C12P21/00,C12R1:125);

CC strandedness: Double;

CC topology: Linear;

CC hypothetical: No;

CC anti-sense: No;

CC \*source: clone=PSAK148;

FH Key location/Qualifiers

FT promoter 193..222

FT RBS /note='SAK promoter'

FT sig\_peptide 301..305

FT 313..393

```
FT      mat_peptide      /note='SAK signal peptide'
FT      FT              394..801
FT      FT              /product='staphylokinase'
FT      FT              /note='SAK'
FT      CDS              313..804
FT      FT              /product='precursor of staphylokinase'.
FEATURES
    source      1..804
                /organism='Staphylococcus aureus'
                /db_xref='taxon:1280'
BASE COUNT      294 a      89 c      148 g      273 t
ORIGIN
alignment_scores:
    Quality: 711.00      Length: 136
    Ratio: 5.228      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 99.265
alignment_block:
US-09-728-670-10 x E02873      ..
Align seg 1/1 to: E02873 from: 1 to: 804
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTy 17
|||||
394 TCAAGTTCATTTCGACAAAGAAATATATAAAAGCGAGTACGCGAGTTA 443
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
444 TTTTGAACCAACAGCGCCGATTTGATGTAAATGTGACTGCGAGTTGATG 493
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::
494 GTAAAGGAATGAATGCTATGCCCTCATTTATGTCGAGTTCTCTATTA 543
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
|||||
544 CCTGGGACTACACTTACAAAAGAAAATGAATACATGTCGAATGGGC 593
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
594 ATTAGATCGACACAGCATATAAGAGTTTAGAGTAGTGAAATTAGATCCA 643
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
644 GCGCAAGATCGAAGTCACTTATATGATAAGATAAGAAAAGAAAGAA 693
101 ThrLysSerPheProIleThrGluLysGlyPheValProAspLeu 117
|||||
694 ACGAAGTCTTCCCTATAACAGAAAAGTTTGTGTCCAGATTATC 743
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
|||||
744 AGAGCATATTAAACCCTGATTCACCTTATTACAAGGTTGTATAG 793
134 luLysLys 136
|||||
794 AAAAGAAA 801
seq_name: gb_ba:SPSAK1
seq_documentation_block:
LOCUS      SPSAK1      1377 bp      DNA      linear      BCT 30-MAR-1995
DEFINITION      Staphylococcus aureus s-phi-C gene for staphylokinase.
ACCESSION      X00127
VERSION      X00127.1 GI:47425
KEYWORDS      kinase; plasminogen activator; signal peptide; staphylokinase.
SOURCE      Staphylococcus aureus.
ORGANISM      Staphylococcus aureus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE      1 (bases 1 to 1377)
```

```
AUTHORS      Sako, T. and Tsuchida, N.
TITLE      Nucleotide sequence of the staphylokinase gene from Staphylococcus
JOURNAL      aureus
MEDLINE      Nucleic Acids Res. 11 (22), 7679-7693 (1983)
84069795
FEATURES
    source      1..1377
                /organism='Staphylococcus aureus'
                /strain='phage s-phi-C'
                /db_xref='taxon:1280'
                /db_xref='GI:758303'
                /note='poss. promoter region'
                /note='poss. promoter region'
                /note='ribosome binding site'
                /codon_start=1
                /transl_table=11
                /product='staphylokinase'
                /protein_id='CAA24957.1'
                /db_xref='GI:758303'
                /translation='MLKRSLLFLTVLLLFSSSITNEVSASSSFDKGYKKGDASY
                FEPTGPLYMNVNTGVGKGNELSPHYVEFPKPGTTLTKETIYVWALDATAYKE
                FRVELDPSAKIEVITYYDKNKKKEETKSFPTTEKGFVVPDLSEHIKNGFNLTIKVVI
                EKK'
sig_peptide      313..393
                /note='putative'
mat_peptide      394..801
                /product='staphylokinase'
BASE COUNT      452 a      184 c      255 g      486 t
ORIGIN
alignment_scores:
    Quality: 711.00      Length: 136
    Ratio: 5.228      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 99.265
alignment_block:
US-09-728-670-10 x SPSAK1      ..
Align seg 1/1 to: SPSAK1 from: 1 to: 1377
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTy 17
|||||
394 TCAAGTTCATTTCGACAAAGAAATATATAAAAGCGAGTACGCGAGTTA 443
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
444 TTTTGAACCAACAGCGCCGATTTGATGTAAATGTGACTGAGATTGATG 493
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::
494 GTAAAGGAATGAATGCTATGCCCTCATTTATGTCGAGTTCTCTATTA 543
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
|||||
544 CCTGGGACTACACTTACAAAAGAAAATGAATACATGTCGAATGGGC 593
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
594 ATTAGATCGACACAGCATATAAGAGTTTAGAGTAGTTGAATTAGATCCA 643
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
644 GCGCAAGATCGAAGTCACTTATATGATAAGATAAGAAAAGAAAGAA 693
101 ThrLysSerPheProIleThrGluLysGlyPheValProAspLeu 117
|||||
694 ACGAAGTCTTCCCTATAACAGAAAAGGTTTGTGTCCAGATTATC 743
```

117 rgluhisilelysasnproglypheasnleuilethrlysvaivalleg 134  
|||||  
744 AGAGCATATTAAACCCTGGATCAACTTAATTACAAGGTTGTATAG 793  
134 lulyslys 136  
|||||  
794 AAAAGAAA 801

seq\_name: gb\_pat:AX073708

seq\_documentation\_block:

LOCUS AX073708 1377 bp DNA linear PAT 06-FEB-2001  
DEFINITION Sequence 1 from Patent WO0104287.  
ACCESSION AX073708  
VERSION AX073708.1 GI:12710128

KEYWORDS

SOURCE

ORGANISM

Staphylococcus aureus.  
Staphylococcus aureus  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Staphylococcus.

REFERENCE 1 (bases 1 to 1377)

AUTHORS Halkier,T., Pedersen,A.H. and Okkels,J.S.

TITLE A method for preparing modified polypeptides

JOURNAL Patent: WO 0104287-A 1 18-JAN-2001;

Maxygen Aps (DK)

FEATURES

source

1. .1377  
/organism="Staphylococcus aureus"  
/db\_xref="taxon:1280"

promotor 193. .222

RBS 301. .305

sig\_peptide 313. .393

BASE COUNT 452 a 184 c 255 g 486 t

ORIGIN

alignment\_scores:

Quality: 711.00

Ratio: 5.228

Percent Similarity: 100.000

Percent Identity: 99.265

alignment\_block:

US-09-728-670-10 x AX073708 ..

Align seg 1/1 to: AX073708 from: 1 to: 1377

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17  
|||||  
394 TCAAGTTCATTGACAAAGAAATATATAAAGGCGATGACGCGAGTTA 443  
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
444 TTTGAAACCAACAGCGCCGATTTGATGTAATGTGACTGAGTTGATG 493  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
:::|||||  
494 GTAAAGGAATGAATGCTATCCCTCATATGTCGAGTTTCCATATAA 543  
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTPrAl 67  
|||||  
544 CCTGGGACTACACTTACAAAGAAATAATGTAATCTATGTGCAATGGGC 593  
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84  
|||||  
594 ATTAGATGCGACAGCATATAAGAGTTTAGAGTAGTTGAATTAGATCAA 643  
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysLysGluGlu 100  
|||||  
644 GCGCAAGATCGAAGTCACTTATTATGATTAAGATAAGAAAAAGAGAA 693  
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117  
|||||  
694 ACGAAGTCTTCCCTATTAACAGAAAAAGTTTGTGTCTCCAGATTATTC 743

117 rgluhisilelysasnproglypheasnleuilethrlysvaivalleg 134  
|||||  
744 AGAGCATATTAAACCCTGGATCAACTTAATTACAAGGTTGTATAG 793  
134 lulyslys 136  
|||||  
794 AAAAGAAA 801

seq\_name: gb\_ba:AP003364

seq\_documentation\_block:

LOCUS AP003364 348650 bp DNA linear BCT 07-FEB-2002  
DEFINITION Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete  
sequence, section 7/9.  
ACCESSION AP003364 BA000017  
VERSION AP003364.2 GI:14247707

KEYWORDS

SOURCE

ORGANISM

Staphylococcus aureus subsp. aureus Mu50 (sub-species:aureus Mu50,  
strain:Mu50) DNA.

Staphylococcus aureus subsp. aureus Mu50

Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Staphylococcus.

REFERENCE 1 (sites)

AUTHORS

Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,  
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,  
Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A.,  
Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C.,  
Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J.,  
Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,  
Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.  
Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus  
Lancet 357 (9264), 1225-1240 (2001)  
21311952  
2 (bases 1 to 348650)

REFERENCE 2 (bases 1 to 348650)

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

On May 29, 2001 this sequence version replaced gi:13873637.  
Location/Qualifiers

1. .348650

/organism="Staphylococcus aureus subsp. aureus Mu50"

/strain="Mu50"

/sub\_species="aureus Mu50"

/db\_xref="taxon:158878"

complement(59. .235)

/gene="SAV1935"

complement(59. .235)

/gene="SAV1935"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAB58097.1"

/db\_xref="GI:14247708"

/translation="MKHLTKIFVLAITLFTIGYVLQATNHSQGIKLLAAIMEMIC  
AFISRSNDRKNSK"

775. .2061

/gene="SAV1936"

775. .2061

/gene="SAV1936"

/note="similar to aspartate transaminase protein"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAB58098.1"

/db\_xref="GI:14247709"

/translation="MNPLAQSLNEQLQOSNATAFTMLSDLCOMFYPKILSQAERK  
STTYNATIGMATNKDGKMFASSLDAMFNDLTPDEIFPYAPPGIEELRLDWQKMLRD



gene  
CDS  
NPESIDNMSLPVITNALTHGLSLVGLFVNQGFILLPEHNMGNKLVENTRNGANL  
QTYPIFDKDHXYTDSLVEALQSYNKDKVIMILNYPNNPFGYPTHKEVTTIVDAIKA  
LADKGTAVAVDDAYYGLFYEDVYQSFLTALSNLNSNAILEPRLDGATKEFFAWGF  
RVGFMTEGTSDDQTKVELAKVGLIRSNISSGPLTQSAVKHVKNNKQFDEIEQN  
IQTLKERYEVTKEVVYADQYHSHWQAYDFNSGYFMAIKVHDVDEELRKHLEKYSIG  
VIALNATDIRIAFSCVEKEDIPIHFDSIAKAIDDLR"  
complement(2340. .4075)  
/gene="truncated-mapw"  
/gene="truncated-mapw"  
/note="SAV1937"  
/codon\_start=1  
/transl\_table=11  
/product="truncated map-w protein"  
/protein\_id="BAB58099.1"  
/db\_xref="GI:14247710"  
/translation="MKKINKSNVKKFPYTIINKFENIVSNEFYFNASKITINDLSIK  
LKSAMANDOGITTKHDIGLAERAVYKVFKNSSKYVDLKTKEYKDERVEKATDIKKVDI  
ELKF"  
CDS  
complement(2645. .4075)  
/gene="truncated-mapw"  
/note="SAV1938"  
/codon\_start=1  
/transl\_table=11  
/product="truncated map-w protein"  
/protein\_id="BAB58100.1"  
/db\_xref="GI:14247711"  
/translation="MKFKSLITTTTALGVIASTGANFNTNEASAAPLDDSSSTLHH  
GHSNIQIPYTTVNGTSQNLISLTFENKQNISYKDIENKVSQVYFNKGISDIDLRL  
SKQAEYTVHFKNGTKRVIDLKSQYITADLINTSDIKAISVNVDPKQPKDKAKANQV  
PYTTVNGTSQNLISLTFENKQNISYKDEGKVSYLESNRGITVDLRLSKQAKYT  
VNFKNGTKKVIDLKSQYITANLINSIDKISININVDTKKHIEKAKRNYQVPYSINL  
GSTNII.SNLSFSNKPWTNYKNLSQKSVLKHDRGISEQDLKAKKAYTVYFKNGG  
KRILQNSKNYANLVHAKDVKRIEITVGTAKADRYVPTIANVNGSTPILSKLK  
ISNKQLISYKYLNDKVKVSKSERGISDLDLFAKAKYTVYFKNGKQVNLKSDIF  
TPNLFSAKDIDKIDIVKQYTKSKKK"  
4418. .4660  
/gene="truncated-hlb"  
4418. .4660  
/gene="truncated-hlb"  
/note="SAV1939"  
/codon\_start=1  
/transl\_table=11  
/product="truncated 6449-6450 beta-hemolysin"  
/protein\_id="BAB58101.1"  
/db\_xref="GI:14247712"  
/translation="MLHLFCCKLYKRSNDGCKTKTSNSLKKVATLALANLLVGALTJ  
NSAKAESKKDDTDLKLVSHNVMLSTVLYPNWRLLT"  
5032. .5211  
/gene="SAV1940"  
5032. .5211  
/gene="SAV1940"  
/note="Bacteriophage phin315"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB58102.1"  
/db\_xref="GI:14247713"  
/translation="MKFKTKCKSNINIFKRSQEIINNSESEKGLGISMILLFLLI  
LEGITACSSSIHFIN"  
5370. .5543  
/gene="SAV1941"  
5370. .5543  
/gene="SAV1941"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB58103.1"  
/db\_xref="GI:14247714"  
/translation="MIGIFVAIYGMKRATYWLMLILFTNIFYFLHFCVLLLLKYIGF  
KLFIYEGCVLLFT"  
complement(5834. .6184)  
/gene="SAV1942"

CDS  
complement(5834. .6184)  
/gene="SAV1942"  
/note="Bacteriophage phin315"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB58104.1"  
/db\_xref="GI:14247715"  
/translation="MKIRKSIAGTLAIVLASPLVTNLDKNEAQAFTSLPTSEYONE  
KLANELKSLDELNVNELATGSLNTYKRTIKISGOKAMYALKSKDFKMSKAYQLO  
KIYNEIDEALSKSY"  
complement(6695. .7162)  
/gene="truncated-lyta"  
complement(6695. .7162)  
/gene="truncated-lyta"  
/note="Bacteriophage phin315  
SAV1943"  
/codon\_start=1  
/transl\_table=11  
/product="truncated amidase"  
/protein\_id="BAB58105.1"  
/db\_xref="GI:14247716"  
/translation="MLAIDILFALHSIHISQHRNITVATPHLKWGCFYWKCVSGIS  
NVKTPAKNEKNSPVAGYTLDDKNVNPYKKEGTYYTVANVKGNNVRDGYSTNSRITGYL  
PNNATIKYDGAYCINGYRMITYIANNGORCYIATGEVDKAGNRISSEFNFSAL"  
complement(7682. .8173)  
/gene="sak"  
complement(7682. .8173)  
/gene="sak"  
/note="Bacteriophage phin315  
SAV1944"  
/codon\_start=1  
/transl\_table=11  
/product="staphylokinase precursor"  
/protein\_id="BAB58106.1"  
/db\_xref="GI:14247717"  
/translation="MLKRSLLFTVLLLLFSSSITNEVSASSSFDKGYKKGDDASY  
FEPTGPYLMVNTGVGDKGNELSPHYVEFPKPGTTLTKEKIEYVEWALDATAKKE  
FRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGVVVDLSEHIKNPGNLLITKVV  
EKK"  
complement(8388. .9119)  
/gene="SAV1945"  
complement(8388. .9119)  
/gene="SAV1945"  
/note="Bacteriophage phin315"  
/codon\_start=1  
/transl\_table=11  
/product="lytic enzyme"  
/protein\_id="BAB58107.1"  
/db\_xref="GI:14247718"  
/translation="MKTYSERARLRWYQGRYIDFDGWYGYQCADLAVDIYWLLEIR  
MWGNAKDAIINDFKNMATVYENTPSPFOIGDVAFTKGIYKQYGHIGLVNGGNTNQ  
FLIEQNYDGNANTPAKLKMDNYGCTHFIRPKYSGSLMKNKITNKVKPRAOKAVGKS  
ASKITVGSKAPYNLKWSKGAFNAKIDGLGATSAFRRYGDNRNTNYRFDVGQAVYAPGTL  
IYVEIIDGWCRIYWNHNENICYE"  
complement(9131. .9385)  
/gene="SAV1946"  
complement(9131. .9385)  
/gene="SAV1946"  
/note="Bacteriophage phin315"  
/codon\_start=1  
/transl\_table=11  
/product="holin homolog"  
/protein\_id="BAB58108.1"  
/db\_xref="GI:14247719"  
/translation="MINWKIRMKOKSEFWAIIISAIFLAQNIKAIGYDIQYYTEQLT

alignment\_scores:  
Quality: 711.00 Length: 136  
Ratio: 5.228 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.265  
alignment\_block:

US-09-728-670-10 x AP003364/rev ..

Align seg 1/1 to reverse of: AP003364 from: 1 to: 348650

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17  
|||||  
8092 TCAAGTTCATTCGACAAAGGAAATATATAAAAGCGATGACCGGAGTTA 8043

17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34  
|||||  
8042 TTTTGAACCAACAGGCCCGTATTGATGTAATGTGACTGAGTTGATG 7993

34 eRLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
:::|||||  
7992 GTAAAGGAATGAATGCTATCCCTCATATTGTCGAGTTTCCATTATAA 7943

51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrAl 67  
|||||  
7942 CCTGGACTACACTACAAAGAAAAATGAAATACTATGTCGAATGGGC 7893

67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84  
|||||  
7892 ATTAGATGCGACAGCATATAAAGAGTTAGACTAGTTGAATTAGATCCAA 7843

84 eRLAlaLysIleGluValThrTyrTyrAspLysAsnLysLysLysGlu 100  
|||||  
7842 GCCCAAGATCGAAGCTCATTATTATGATAGAATAAGAAAAAGAGAA 7793

101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117  
|||||  
7792 ACGAAGTCTTCCCTATTAACAGAAAAAGGTTTGTGTCCAGATTATTC 7743

117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134  
|||||  
7742 AGAGCATATTAAAAACCTGATTCAACTTAATTACAAAGGTTGTTATAG 7693

134 LuLysLys 136  
|||||  
7692 AAAAGAAA 7685

seq\_name: gb\_ba:AB033232

seq\_documentation\_block:

LOCUS AB033232 15600 bp DNA linear BCT 16-MAY-2000

DEFINITION Staphylococcus aureus genes for kdpA homolog, kdpC homolog, hypothetical protein, staphylokinase, tagA homolog, tagH homolog, tagG homolog, complete cds.

ACCESSION AB033232

VERSION AB033232.1 GI:7670319

KEYWORDS hypothetical protein; TagG homolog; TagH homolog; TagA homolog; staphylokinase; kdpB homolog; kdpA homolog.

SOURCE Staphylococcus aureus

ORGANISM Staphylococcus aureus

REFERENCE 1 (sites)

AUTHORS Horii,T., Yokoyama,K., Barua,S., Odagiri,T., Futamura,N., Hasegawa,T. and Ohta,M.

TITLE The staphylokinase gene is located in the structural gene encoding N-acetylmuramyl-L-alanine amidase in methicillin-resistant staphylococcus aureus

JOURNAL FEMS Microbiol. Lett. 185 (2), 221-224 (2000)

MEDLINE 20219005

REFERENCE 2 (bases 1 to 15600)

AUTHORS Horii,T., Yokoyama,K., Barua,S., Odagiri,T., Futamura,N. and Ohta,M.

TITLE Direct Submission

JOURNAL Submitted (04-OCT-1999) Toshinobu Horii, Nagoya University School of Medicine, Department of Bacteriology; 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan

FEATURES (E-mail:horii@etsuru.med.nagoya-u.ac.jp, Tel:052-744-2101, Fax:052-732-1519)

Location/Qualifiers

source

1. .15600

/organism="Staphylococcus aureus"

/db\_xref="taxon:1280"

333. .926

/codon\_start=1

/transl\_table=1

/product="kdpA homolog"

/protein\_id="BAA95007.1"

/db\_xref="GI:7670320"

/translation="MVLMLNAYRGEGVGMLNMLIYVMLTVEICSLMIGKTPSYLGM KIEGKEMKIALSFLVHPLLILVFSALAFIVPGASDALNPQFHGVSOVLVEFTSSA NNGSGFEGLDNTVEFWNISTGIVMLARYIPVLQILVSSLVNKTTYQHTQDYPIN NLFSSVLIIFILLSGLTFPLDMLGRIGEQLLHA"

945. .1910

/codon\_start=1

/transl\_table=1

/product="kdpB homolog"

/protein\_id="BAA95008.1"

/db\_xref="GI:7670321"

/translation="MAETTKIFESHLYKQALKDSVLKLYPYMIKNPIMFVEVGMLL ALGLIYPLDFHQESVSRLYVFSFIILLTLVFANFSELAEGRKAQANALROTQ EMKARRIKODGSYEMIDASDLKCHIVRVATGEOIPNDGKVIKGLATVDESATGESA PVIKESGDFDNVIGTSSVADMLEVEITSEPGHSFLDMIGLVEGATRKTPNETAL FTLMTLTIFLIVILTMYPFLAKFNLSIAMLIALVCLIPITIGLLSAIGIAGM DRYQFNILAKSGRSVETCGDVNVLLDKTGTITTAIVQMRLFR"

2980. .3537

/codon\_start=1

/transl\_table=1

/product="kdpC homolog"

/protein\_id="BAA95009.1"

/db\_xref="GI:7670322"

/translation="MQTIRKSLGLVILMFVLCGFIFPLTYALGOVLPPEQANGSLVK ODGKVIKGLIGQWTEPKYFHGRI SAVNYNMNANEVKSGSPASGSGSNYGNSEPK KRVQETIKQEGKISSDAVTASGSLDPDITVDNAKQVYRIAKERINIDASKINHLID ENKQASPMADDDYVNVLLKINTLIDKL"

5236. .5490

/codon\_start=1

/transl\_table=1

/product="hypothetical protein"

/protein\_id="BAA95010.1"

/db\_xref="GI:7670323"

/translation="MINWKIRMKQKSEFWAIIISAIFLFAQNIKAIGYDIQVYTEQLT DGLNALIGFPVLTGVITQPTTKIGIDRHQALEEPRRKY"

6448. .6939

/codon\_start=1

/transl\_table=1

/product="staphylokinase"

/protein\_id="BAA95011.1"

/db\_xref="GI:7670324"

/translation="MLKRGLFLTYVLLLLSFSSSITNEVSASSSFDKGYKKGDASY FEPTGPLYMNVNTGVDGKNELSPHYVEPIKPGTTLTRKEIYYVEWALDATAYKE FRVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDISEHIKNPGFNLTIVYII EKK"

9120. .9815

/codon\_start=1

/transl\_table=1

/product="tagA homolog"

/protein\_id="BAA95012.1"

/db\_xref="GI:7670325"

/translation="MLQMVENIKTFEFAQSTNNLFIYVANDIYVYATTQAYLELIN QASYIVADGCVVKASHRKQPLNHRIPGIELMDECLKIAHYNHQVFLGATNEVE AAQYALQQRYPNISFAHHGYIDLEDTVVKRIKLFKPDYIFVGMGFPKQEEWIMTHE NOFESTVMGVGGSLEVFAGAKKRADYIFRKLNIEMIYRALIDWKIRGLKSIPLEMY KIAKAKRIKKAK"

complement(9876. .10670)

/codon\_start=1

/transl\_table=1

/product="tagH homolog"

/protein\_id="BAA95013.1"

/db\_xref="GI:7670326"

/translation="MNVSVNIKVYKRYRIYRTNKERMKDALIPKHNKTFALDDIS LKAYEGDIVIGLVINGSGKSTLSNIGGSLSPYGVKVDNRGEVSVAIASAGLSQLTG IENIEFKMLCMGFKRKEIKAMTPKIIEFSELGEFIYQPVKYYSSGMRAKLGSINTY

```
NPDLIVIDEALSVGDQTFQAKCLDKIYEKQNKTIFFVSHNLQVROFCKTIANIES
GKLKDGELDDVLPKYEAFNLNFKRSKAEOKEFRNKLDKARFVIK"
11017.12066
/codon_start=1
/transl_table=11
/product="Tagg homolog"
/protein_id="BAA95014.1"
/db_xref="GI:7670327"
/translation="MSAIGTVEKHEVKNFYLIQRLAQFYKINHSNYLGVAMELINP
VMOIMYVMVVEGLIRSNAPIHGPEVYWLVLVGISMWEFINOGLIEGKAITQKENVY
SKMNEPLSIPTIYIVTSREYGLGLLLVIACMFTGIYPSIHIIQLIYVPCFELT
ASVTLSTSLGLVLRDQMLMQAILRLIFYFSPILWPKHGISGLIHEMKYNPYVF
IAESYRAAILLHEWYFMDPSALVYLVAPTNOTNGSEADQIPSGYTIILATGTPDGVA
NTTIRQDYYVFIPVGVKQIRAVVYINKVVASNMSNAVITLLPDDIPTINNPGVINA
KYRGDEVNFTMGVC"
BASE COUNT      5264 a      2419 c      2914 g      5003 t
ORIGIN

alignment_scores:
      Quality:  710.00      Length:   136
      Ratio:    5.221       Gaps:    0
Percent Similarity: 100.000   Percent Identity: 98.529

alignment_block:
US-09-728-670-10 x AB033232 ..
Align seg 1/1 to: AB033232 from: 1 to: 15600

1 SerSerSerPheAspLysGlyLysTyrrLysGlyAspAspAlaSerTy 17
|||||
6529 TCAGATTCATTCGACAAAGAAATATATAAAAGCGCATGACCGAGTTA 6578
17 rPheGluProThrGlyProTyrrLeuMetValAsnValThrGlyValAsp 34
|||||
6579 TTTTGAACCAACAGAGCCCGCTATTGTGATGTAATGTGACTGAGTTGATG 6628
34 erLysGlyAsnGluLeuLeuSerProHisTyrrValGluPheProIleLys 50
:::
6629 GTAAAGGAATGAATGCTATCCCTCATTTATGTCGAGTTCTATTAA 6678
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrrTyrrValGluTrpAl 67
|||||
6679 CCTGGGACTACACTTACAAAGAAATGAATGATGTCGAAATGGGC 6728
67 aleuAspAlaThrAlaTyrrLysGluPheArgValGluLeuAspPro 84
|||||
6729 ATTAGATGCGACAGCATATAAGAGTTTAGAGTAGTGAATTAGATCCAA 6778
84 eraLalysIleGluValThrTyrrTyrrAspLysAsnLysLysGluGlu 100
|||||
6779 GCGCAAGATCGAAGTCACTTATATGATAAGAAATAGCAAAAGAGAA 6828
101 ThrLysSerPheProIleThrGluLysGlyPheValProAspLeu 117
|||||
6829 ACCGAAGTCTTTCCTATACAGAAAAAGTTTGTGTCCTCCAGATTATC 6878
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
6879 AGAGCATATTAAACCCCTGATTCACACTTAATTACAAAGCTTATTATAG 6928
134 luLysLys 136
|||||
6929 AAAAGAAA 6936
seq_name: gb_ba:AF332619
seq_documentation_block:
LOCUS      AF332619              727 bp    DNA      linear      BCT 21-FEB-2001
DEFINITION  Staphylococcus aureus staphylokinase SakXH gene, complete cds.
ACCESSION  AF332619
VERSION    AF332619.1    GI:13022096
KEYWORDS
```

```
SOURCE      Staphylococcus aureus.
ORGANISM    Staphylococcus aureus
REFERENCE   1 (bases 1 to 727)
AUTHORS     Wei, W., Xiang, H. and Tan, H.
TITLE       Direct Submission
JOURNAL     Submitted (28-DEC-2000) Genetics, Institute of Microbiology,
            Chinese Academy of Sciences, 13, Beiyitiao, Zhongguancun, Beijing
            100080, P.R.China
FEATURES
  source
    1..727
    /organism="Staphylococcus aureus"
    /db_xref="taxon:1280"
    224..228
    236..727
    /function="catalyzes plasminogen to plasmin"
    /note="protein kinase; similar to staphylokinase variants
    sak42D, sakSTAR and sakC"
    /codon_start=1
    /transl_table=11
    /product="staphylokinase SakXH"
    /protein_id="AK11636.1"
    /db_xref="GI:13022097"
    FEPTGYLMVNVTVGVDGKNELSPRYVEFPKPGTTLTKETIYVEMALDATAYKE
    FRVVELDPSAKIEVYYDKNKKKEETKSFITEKGFVVPDLSEHKNPENLITKVI
    EKK"
BASE COUNT      269 a      80 c      132 g      246 t
ORIGIN

alignment_scores:
      Quality:  703.00      Length:   136
      Ratio:    5.169       Gaps:    0
Percent Similarity: 100.000   Percent Identity: 98.529

alignment_block:
US-09-728-670-10 x AF332619 ..
Align seg 1/1 to: AF332619 from: 1 to: 727

1 SerSerSerPheAspLysGlyLysTyrrLysGlyAspAspAlaSerTy 17
|||||
317 TCAGATTCATTCGACAAAGAAATATATAAAAGCGCATGACCGAGTTA 366
17 rPheGluProThrGlyProTyrrLeuMetValAsnValThrGlyValAsp 34
|||||
367 TTTTGAACCAACAGAGCCCGCTATTGTGATGTAATGTGACTGAGTTGATG 416
34 erLysGlyAsnGluLeuLeuSerProHisTyrrValGluPheProIleLys 50
:::
417 GTAAAGGAATGAATGCTATCCCTGTTATGTCGAGTTCTATTAA 466
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrrTyrrValGluTrpAl 67
|||||
467 CCTGGGACTACACTTACAAAGAAATGAATGATGTCGAAATGGGC 516
67 aleuAspAlaThrAlaTyrrLysGluPheArgValGluLeuAspPro 84
|||||
517 ATTAGATGCGACAGCATATAAGAGTTTAGAGTAGTGAATGATTCCAA 566
84 eraLalysIleGluValThrTyrrTyrrAspLysAsnLysLysGluGlu 100
|||||
567 GCGCAAGATCGAAGTCACTTATATGATAAGAAATAGCAAAAGAGAA 616
617 ACGAAGTCTTTCCTATACAGAAAAAGTTTGTGTCCTCCAGATTATC 666
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
667 AGAGCATATTAAACCCCTGATTCACACTTAATTACAAAGCTTATTATAG 716
```

134 lLyslys 136  
|||||||  
717 AAAAGAA 724

seq\_name: gb\_ba:SAU77328

seq\_documentation\_block:

LOCUS SAU77328 411 bp DNA linear BCT 14-JUN-1999  
DEFINITION Staphylococcus aureus staphylokinase gene, partial cds.  
ACCESSION U77328  
VERSION U77328.1 GI:2605637

KEYWORDS

SOURCE Staphylococcus aureus.

ORGANISM

Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Staphylococcus.

REFERENCE 1 (bases 1 to 411)

AUTHORS Kim,S.H., Chun,H.S., Han,M.H., Park,N.Y. and Suk,K.  
TITLE A novel variant of staphylokinase gene from Staphylococcus aureus  
JOURNAL ATCC 29213  
MEDLINE Thromb. Res. 87 (4), 387-395 (1997)  
PUBMED 9271816

REFERENCE 2 (bases 1 to 411)

AUTHORS Chun,H.S., Suk,K. and Kim,S.H.  
TITLE Direct Submission  
JOURNAL Submitted (04-NOV-1996) Protein Function R.U., KRIBB, KIST, PO Box 115, Yusong, Taejeon 305-600, Korea

FEATURES

source 1. 411  
/organism="Staphylococcus aureus"  
/strain="ATCC 29213"  
/db\_xref="taxon:1280"

CDS  
1. 411  
/codon\_start=1  
/transl\_table=11  
/product="staphylokinase"  
/protein\_id="AAB84174.1"  
/db\_xref="GI:2605638"

translation="SSSFDKGGKKGGDASYFEPTGPLYMNVNTGVEKENELSPHY  
VEFPKRGITLTKEKIEYVEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETK  
SEFITEKGEVVPDLSEHIKNPGENLITKVIEKK"

BASE COUNT 162 a 59 c 81 g 109 t  
ORIGIN

alignment\_scores:  
Quality: 699.00 Length: 136  
Ratio: 5.178 Gaps: 0  
Percent Similarity: 99.265 Percent Identity: 97.794

alignment\_block:  
US-09-728-670-10 x SAU77328 ..

Align seg 1/1 to: SAU77328 from: 1 to: 411

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTy 17  
|||||  
1 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGAGTTA 50

17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
51 TTTTGAACCAACACAGCGCCGTATTGTGATGCTAAATGTGACTGAGTTGAGG 100

34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
:::|  
101 GTAAAGAAATGAATGCTATCCCTCATATATGTCGAGTTCTTAA 150

51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67  
|||||  
151 CCTGGGACTACACTTACAAAAAGAAAAATGAATACTATGTGCAATGGGC 200

67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84  
|||||  
201 ATTAGATGCGACAGCATATAAAGAGTTTAGAGTAGTTGAATTAGATCCAA 250

84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100  
|||||  
251 GCGCAAGATCGAAGTCACTTATGATGAATGAAGAAAAAGAGAA 300

101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117  
|||||  
301 ACGAAGCTTTCCCTATTAACAGAAAAAGTTTGTGTGCCAGATTATC 350

117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134  
|||||  
351 AGAGCATATTAAAAACCCTGGAATTCAACTTAATTAACAAAGTTGTATAG 400

134 lLyslys 136  
|||||||  
401 AAAAGAA 408

seq\_name: gb\_pat:A17529

seq\_documentation\_block:

LOCUS A17529 414 bp DNA linear PAT 12-APR-1994  
DEFINITION Bacteriophage 42D recombinant DNA for plasmidogen activator SAK.  
ACCESSION A17529  
VERSION A17529.1 GI:490010

KEYWORDS staphylokinase.  
SOURCE Bacteriophage 42D.  
ORGANISM Bacteriophage 42D.

Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
Lambda phage group.  
REFERENCE 1 (bases 1 to 414)

AUTHORS  
TITLE EXPRESSION SIGNAL-PEPTIDE-FREE STAPHYLOKINASES  
JOURNAL Patent: WO 9313209-A 1 08-JUL-1993;  
FEATURES Location/Qualifiers

source 1. 414  
/organism="Bacteriophage 42D"  
/specific\_host="Staphylococcus aureus"  
/db\_xref="taxon:10715"

CDS  
1. 414  
/codon\_start=1  
/transl\_table=11  
/product="staphylokinase SAK42D"  
/protein\_id="CAA01335.1"

translation="MSSSFDKGGKKGGDASYFEPTGPLYMNVNTGVDGRNELLSPR  
YVEFPKRGITLTKEKIEYVEWALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEET  
KSFPIERKGEVVPDLSEHIKNPGENLITKVIEKK"

mat\_peptide 4. 411  
/product="staphylokinase SAK42D"

BASE COUNT 161 a 59 c 82 g 111 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 695.00 Length: 136  
Ratio: 5.148 Gaps: 0  
Percent Similarity: 99.265 Percent Identity: 97.794

alignment\_block:  
US-09-728-670-10 x A17529 ..

Align seg 1/1 to: A17529 from: 1 to: 414

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTy 17  
|||||  
4 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGAGTTA 53

17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
54 TTTTGAACCAACAGCGCCGTATTGTGATGCTAAATGTGACTGAGTTGATG 103





```

124 TCAAGTTCATTCGACAAGAAATATAAAAAAGCGATGACGGAGTTA 173
17  rphEgluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
   |||||||
174 TTTTGAACCAACAGAGCCCGCTATTGATGCTAAATGTGACTGGAGTTGATG 223
34  erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
   :||| |||||||
224 GTAAAGAAATGAATGCTATCCCTCGTTATGTCGAGTTTCTATTAA 273
51  ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrpa 67
   |||||||
274 CTGGGACTACACTTACAAAGAAATGAATGACTATGTCGAATGGGC 323
67  aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
   |||||||
324 ATTAGATGCCAGACATATAAAGAGTTTAGAGTAGTTGAATGAAACCA 373
84  erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
   |||||||
374 GCGCAAGATCGAAGTCACTTATATGATAAGATAAGAAAAAGACAA 423
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
   |||||||
424 ACGAAGTCTTCCCTATACAGAAAAAGGTTTGTGTGCCAGATTATC 473
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
   |||||||
474 AGAGCATATTAAAAACCCTGATTCACCTTAATTACAAAGGTGTATTAG 523
134 LuLysLys 136
   |||||||
524 AAAAGAAA 531

```

seq\_name: gb\_pat:AX247537

seq\_documentation\_block:

```

LOCUS      AX247537          540 bp      DNA      linear      PAT 28-SEP-2001
DEFINITION Sequence 13 from Patent WO0166776.
ACCESSION  AX247537
VERSION    AX247537.1  GI:15862230
KEYWORDS
SOURCE     .
           Escherichia coli.
ORGANISM   Escherichia coli
           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE  1 (bases 1 to 540)
AUTHORS   Holschen,C., Gumpert,J., Kujau,J.M., Fritzsche,C., Elske,G.,
           Fahner,T.B., Sieben,S. and Mueller,H.P.
TITLE      Novel 1-form bacterial strains, method for producing same and the
           use thereof for producing gene products
JOURNAL    Patent: WO 0166776-A 13 13-SEP-2001;
           INSTITUT FUER MOLEKULARE BIOLOGIE E.V. (DE)
FEATURES   location/Qualifiers
           1..540
           /organism="Escherichia coli"
           /db_xref="taxon:562"
CDS        1..540
           /note="LacyH1-Sak"
           /codon_start=1
           /transl_table=11
           /protein_id="CAC88585.1"
           /db_xref="GI:15862231"
           /translation="MYLKTNTNFMFGLFEFFYFPLMGAYEPFPIWLHDINHISLOS
           SSFDKGYKKGDASYFEPTGYPLMVNTGVGDKNELLSPRYVEFPKPGTTLTKK
           IEYVEMALDATAKERVVELDPSAKIEVTYYDKNKKKEETKSPITEKGFVVPDLS
           EHIKNPGFNLITKVVIKK"

```

```

BASE COUNT      192 a      87 c      99 g      162 t
ORIGIN

```

alignment\_scores: Quality: 695.00 Length: 136

Ratio: 5.148 Gaps: 0  
Percent Similarity: 99.265 Percent Identity: 97.794

alignment\_block:

US-09-728-670-10 x AX247537 ..

Align seg 1/1 to: AX247537 from: 1 to: 540

```

1 SerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTy 17
   |||||||
130 TCAAGTTCATTCGACAAGAAATATAAAAAAGCGATGACGGAGTTA 179
17  rphEgluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
   |||||||
180 TTTTGAACCAACAGCCCGCTATTGATGCTAAATGTGACTGGAGTTGATG 229
34  erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
   :||| |||||||
230 GTAAAGAAATGAATGCTATCCCTCGTTATGTCGAGTTTCTATTAA 279
51  ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrpa 67
   |||||||
280 CTGGGACTACACTTACAAAGAAAAATGAATGACTATGTCGAATGGGC 329
67  aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
   |||||||
330 ATTAGATGCCAGACATATAAAGAGTTTAGAGTAGTTGAATTAGATCAA 379
84  erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
   |||||||
380 GCGCAAGATCGAAGTCACTTATATGATAAGATAAGAAAAAGACAA 429
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
   |||||||
430 ACGAAGTCTTCCCTATACAGAAAAAGGTTTGTGTGCCAGATTATC 479
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
   |||||||
480 AGAGCATATTAAAAACCCTGATTCACCTTAATTACAAAGGTGTATTAG 529
134 LuLysLys 136
   |||||||
530 AAAAGAAA 537

```

seq\_name: gb\_pat:AX247541

seq\_documentation\_block:

```

LOCUS      AX247541          639 bp      DNA      linear      PAT 28-SEP-2001
DEFINITION Sequence 17 from Patent WO0166776.
ACCESSION  AX247541
VERSION    AX247541.1  GI:15862234
KEYWORDS
SOURCE     .
           Escherichia coli.
ORGANISM   Escherichia coli
           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE  1 (bases 1 to 639)
AUTHORS   Holschen,C., Gumpert,J., Kujau,J.M., Fritzsche,C., Elske,G.,
           Fahner,T.B., Sieben,S. and Mueller,H.P.
TITLE      Novel 1-form bacterial strains, method for producing same and the
           use thereof for producing gene products
JOURNAL    Patent: WO 0166776-A 17 13-SEP-2001;
           INSTITUT FUER MOLEKULARE BIOLOGIE E.V. (DE)
FEATURES   location/Qualifiers
           1..639
           /organism="Escherichia coli"
           /db_xref="taxon:562"
CDS        1..639
           /note="SecAH1-Sak"
           /codon_start=1
           /transl_table=11
           /protein_id="CAC88587.1"
           /db_xref="GI:15862235"

```

/translation-"MAKOPGLDFQSAKGLGELKRRLLFVIGALIVRIGSFIPPGI  
DAAVLAKLEBOQRCTIEMFMNSGALSRLQSSFDKGYKGGDASYEPTGPYLM  
VNVTVGDKRNELLSPRVVEFPKPGTTLTKKIEIYVEMALDATAYKEPRVELDPS  
AKIEVTYDDKNKKKEETKSFITEKGFVVDLSEHIKNPGFNLTIKVIEKK"  
BASE COUNT 209 a 110 c 141 g 179 t  
ORIGIN

alignment\_scores: Quality: 695.00 Length: 136  
Ratio: 5.148 Gaps: 0  
Percent Similarity: 99.265 Percent Identity: 97.794

alignment\_block:  
US-09-728-670-10 x AX247541 ..

Align seg 1/1 to: AX247541 from: 1 to: 639

```
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTy 17
|||||
229 TCAAGTTCATTCGACAAAGGAAATATATAAAAGCGATGACGCGAGTTA 278
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
279 TTTTGAACCAACAGCGCCGCTATTGATGTAATGTGACTGCGAGTTGATG 328
34 erLysGlyAsnGluLeuSerProHisTyrValGluPheProIleLys 50
:::
329 GTAAAGAAATGAATGCTATCCCTCGTTATGTCGAGTTCTCTATTAAA 378
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrAl 67
|||||
379 CCTGGGACTACACTTACAAAAGAAAATGAATACTATGTCGAATGGGC 428
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
429 ATTAGATGCGACAGCATATTAAGAGTTTAAAGAGTTGAATTAGATCCAA 478
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
479 GCGCAAGATCGAAGTCACTTATATGATAAGATAAGAAAAGAGAGAA 528
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
|||||
529 ACGAAGCTTTCCTATTAACAGAAAAGGTTTGTGTGCCAGATTATC 578
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
579 AGAGCATATTAAAAACCCCTGATCAACTTAATTACAAAGGTTGTATAG 628
134 LuLysLys 136
|||||
629 AAAAGAAA 636
```

seq\_name: gb\_pat:AX247539

seq\_documentation\_block:

LOCUS AX247539 720 bp DNA linear PAT 28-SEP-2001  
DEFINITION Sequence 15 from Patent WO0166776.  
ACCESSION AX247539  
VERSION AX247539.1 GI:15862232  
KEYWORDS  
SOURCE Escherichia coli.  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.

REFERENCE 1 (bases 1 to 720)

AUTHORS Holtschen,C., Gumpert,J., Kujau,J.M., Fritsche,C., Eliske,G.,

Fahner,B., Sieben,S. and Mueller,H.P.

TITLE Novel 1-form bacterial strains, method for producing same and the

use thereof for producing gene products

JOURNAL Patent: WO 0166776-A 15 13-SEP-2001;

INSTITUT FUER MOLEKULARE BIOLOGIE E.V. (DE)

FEATURES Location/Qualifiers

source

1..720

/organism="Escherichia coli"

/db\_xref="taxon:562"

1..720

/note="LacYH1-3-Sak"

/codon\_start=1

/transl\_table=11

/protein\_id="CAC88586.1"

/db\_xref="GI:15862233"

/translation="MYLKNTNFMFGLFFFFVFIMGAYFPFPINLHDINHISKSD  
TGIIFPAISLFLFQPLFLSDKLGIRKYLMLVTFAPFIFGLLOYL  
QSSSFDKGYKKGDASYFEPTGPYLMVNVTVGDKRNELLSPRVVEFPKPGTTLTK  
EKIEYVEMALDATAYKEPRVELDPSAKIEVTYDDKNKKKEETKSFITEKGFVVPD  
LSEHIKNPGFNLTIKVIEKK"

BASE COUNT 226 a 127 c 136 g 231 t  
ORIGIN

alignment\_scores: Quality: 695.00 Length: 136  
Ratio: 5.148 Gaps: 0  
Percent Similarity: 99.265 Percent Identity: 97.794

alignment\_block:  
US-09-728-670-10 x AX247539 ..

Align seg 1/1 to: AX247539 from: 1 to: 720

```
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTy 17
|||||
310 TCAAGTTCATTCGACAAAGGAAATATATAAAAGCGATGACGCGAGTTA 359
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
360 TTTTGAACCAACAGCGCCGCTATTGATGTAATGTGACTGCGAGTTGATG 409
34 erLysGlyAsnGluLeuSerProHisTyrValGluPheProIleLys 50
:::
410 GTAAAGAAATGAATGCTATCCCTCGTTATGTCGAGTTCTCTATTAAA 459
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrAl 67
|||||
460 CCTGGGACTACACTTACAAAAGAAAATGAATACTATGTCGAATGGGC 509
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
510 ATTAGATGCGACAGCATATTAAGAGTTTAAAGAGTTTGAATTAGATCCAA 559
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
560 GCGCAAGATCGAAGTCACTTATTAAGAGTTTGAATTAGATCCAA 609
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
|||||
610 ACGAAGCTTTCCTATTAACAGAAAAGGTTTGTGTGCCAGATTATC 659
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
660 AGAGCATATTAAAAACCCCTGATCAACTTAATTACAAAGGTTGTATAG 709
134 LuLysLys 136
|||||
710 AAAAGAAA 717
```

seq\_name: gb\_pat:AX247543

seq\_documentation\_block:

LOCUS AX247543 876 bp DNA linear PAT 28-SEP-2001  
DEFINITION Sequence 19 from Patent WO0166776.  
ACCESSION AX247543  
VERSION AX247543.1 GI:15862236  
KEYWORDS

```

SOURCE      Escherichia coli.
ORGANISM    Escherichia coli
Bacteria:  Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE   1 (bases 1 to 876)
AUTHORS    Holscher,C., Gumber,J., Kujau,J.M., Fritsche,C., Elske,G.,
Fahner,B., Steben,S. and Mueller,H.P.
TITLE      Novel 1-form bacterial strains, method for producing same and the
JOURNAL    use thereof for producing gene products
            Patent: WO 0166776-A 19 13-SEP-2001;
            INSTITUT FUER MOLEKULARE BIOLOGIE E.V. (DE)
FEATURES
  source
    1..876
    /organism="Escherichia coli"
    /db_xref="taxon:562"
    1..876
    /note="SecYH1-3-Sak"
    /codon_start=1
    /transl_table=11
    /protein_id="CAC88588.1"
    /db_xref="GI:1586237"
    /translation="MAQDPELDQSAKGGELKRLFLVIGALIVERIGSFIPIGI
    DAAVLAKLLEQRTIEMFMESGALSASISFALGIMPYISASIIQLTVHPPL
    AEIKGESGRKISQYTRIGTLVLAIFOSIGIATGLPNMGVLINPGLQSSFD
    KGKYYKGDASYFEPTGYLMVNTGVDGRNELSPRYVEPIKPGTTLTKIEIY
    VEMALDATAYKEFRVVELDPSAKIEVTYYDKNKKKEETKSFITEKGFVPLSEHIK
    NPGFNLTQVIEKK"
BASE COUNT  259 a      168 c      207 g      242 t
ORIGIN
alignment_scores:
  Quality:   695.00      Length:   136
  Ratio:     5.148      Gaps:     0
  Percent Similarity:  99.265      Percent Identity:  97.794
alignment_block:
US-09-728-670-10 x AX247543 ..
Align seg 1/1 to: AX247543 from: 1 to: 876
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
466 TCAGTTCATTCGACAAAGAAATATATAAAGCGCATGACGCGAGTTA 515
17 rphegluprothrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
516 TTTGAACCAACAGGCCGATTTGATGTTAATGTGACTGAGTTGATG 565
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
::|||
566 GTAAAGAAATGAAATGCTATCCCTCGTATGTCGAGTTTCCCTATTAA 615
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrAl 67
|||||
616 CTGGGACTACACTTACAAAAAATAATGATACTATGTCGAATGGGC 665
67 aleuAspAlaThrAlaTyrLysGluPheArgValAlaGluLeuAspPro 84
|||||
666 ATTAGATCGACAGCATATAAGAGTTTAGAGTAGTTGAATTAGATCCAA 715
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
716 GCGCAAGATCGAAGTCACCTTATTATGATAAGATAAGAAAAAGAGAA 765
101 ThrLysSerPheProIleThrGluLysGlyPheValAlaProAspLeuSe 117
|||||
766 ACGAAGTCTTCCCTATAACAGAAAAAGTTTGTGTGCCAGATTATC 815
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
816 AGAGCATATTAAAAACCTGGATTCACACTTAATTACAAAGGTGTATAG 865

```

```

134 LuLysLys 136
|||||
866 AAAAGAAA 873
seq_name: gb_ph:P42STK
seq_documentation_block:
LOCUS      P42STK
DEFINITION Bacteriophage 42D (clone PDB17) (from Staphylococcus aureus)
ACCESSION M57455
VERSION    M57455.1 GI:215344
KEYWORDS   staphylokinase.
SOURCE     Bacteriophage 42D (clone: PDB17) DNA.
ORGANISM   Bacteriophage 42D
            Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
            lambda phage group.
REFERENCE  1 (bases 1 to 993)
AUTHORS    Behnke,D., Gerlach,D. and Kraft,R.
TITLE      Cloning, sequencing, and expression in E.coli and B. subtilis of a
JOURNAL    staphylokinase gene
            (in) Chaioupa,J. and Krumphanz,V. (Eds.);
            EXTRACELLULAR ENZYMES OF MICROORGANISMS: 29-38;
            Plenum Press, New York (1987)
FEATURES
  source
    1..993
    /organism="Bacteriophage 42D"
    /specific_host="Staphylococcus aureus"
    /db_xref="taxon:10715"
    /clone="PDB17"
    194..199
    -35_signal
    -10_signal
    CDS
    314..805
    /codon_start=1
    /transl_table=11
    /product="staphylokinase"
    /protein_id="AAA98206.1"
    /db_xref="GI:215345"
    /translation="MLKRSLLFLVLLLFSSSITNEVSASSSFDKGYKKGDASY
    FEPTGYLMVNTGVDGRNELSPRYVEPIKPGTTLTKIEIYVEMALDATAYKE
    FRVVELDPSAKIEVTYYDKNKKKEETKSFITEKGFVPLSEHIKPNPGFNLTQVI
    EKK"
    902..940
    terminator
    /note="putative"
    956..975
    gene
    /gene="staphylokinase"
    956..975
    terminator
    /gene="staphylokinase"
    /note="putative"
BASE COUNT  356 a      118 c      183 g      336 t
ORIGIN
alignment_scores:
  Quality:   695.00      Length:   136
  Ratio:     5.148      Gaps:     0
  Percent Similarity:  99.265      Percent Identity:  97.794
alignment_block:
US-09-728-670-10 x P42STK ..
Align seg 1/1 to: P42STK from: 1 to: 993
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
395 TCAAGTTCATTCGACAAAGAAATATATAAAGCGCATGACGCGAGTTA 444
17 rphegluprothrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
445 TTTGAACCAACAGGCCGATTTGATGTTAATGTGACTGAGTTGATG 494
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
::|||

```



495 GTAAAGAATGTAATGCTATCCCTCGTTATGTCGAGTTCTATTAA 544  
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67  
|||||  
545 CCTGGAGCTACACTTACAAAAGAAAAAATTGAATACTATGTCGAATGGC 594  
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84  
|||||  
595 ATTAGATCGACAGCATATAAGAGTTTAGAGTAGTTGAATTAGATCCAA 644  
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100  
|||||  
645 GCGCAAGATCGAAGTCACTTATGATAAGAATAAGAAAAAGAAGAA 694  
101 ThrLysSerPheProIleThrGluLysGlyPheValProAspLeu 117  
|||||  
695 ACGAAGTCTTCCCTATACAGAAAAAGGTTTGTGTCCGAGTTTATC 744  
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134  
|||||  
745 AGAGCATATTAAAAACCCTGGATCACTTAATTACAAGGTTGTATAG 794  
134 luLysLys 136  
|||||  
795 AAAAGAAA 802

seq\_name: gb\_ba:SASAK42D

seq\_documentation\_block:

LOCUS SASAK42D 1023 bp DNA linear BCT 30-MAR-1995  
DEFINITION Staphylococcus aureus phage 42D for staphylokinase.  
ACCESSION X06603  
VERSION X06603.1 GI:46676  
KEYWORDS plasmidogen activator; sak42D gene; staphylokinase.  
SOURCE Staphylococcus aureus.  
ORGANISM Staphylococcus aureus  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Staphylococcus.  
REFERENCE 1 (bases 31 to 1023)  
AUTHORS Behnke, D. and Gerlach, D.  
TITLE Cloning and expression in Escherichia coli, Bacillus subtilis, and Streptococcus sanguis of a gene for staphylokinase--a bacterial plasmidogen activator  
JOURNAL Mol. Gen. Genet. 210 (3), 528-534 (1987)  
MEDLINE 88121731  
REFERENCE 2 (bases 1 to 30)  
AUTHORS Behnke, D.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUL-1988)  
COMMENT Data kindly reviewed (09-JUL-1988) by Behnke D.  
FEATURES  
location/Qualifiers  
1..1023  
/organism="Staphylococcus aureus"  
/db\_xref="taxon:1280"  
224..229  
/note="35 region"  
248..253  
/note="10 region"  
332..336  
/note="Shine-Dalgarno sequence"  
344..835  
/codon\_start=1  
/transl\_table=11  
/product="sak42D staphylokinase"  
/protein\_id="CAA29822.1"  
/db\_xref="GI:758275"  
/db\_xref="SWISS-PROT:P15240"  
/translation="MKRSLLFLTVLLLFSSSITNEVSASSSFDKGYKKGDDASY  
FEPTGYLMVNTGVGCKRNLSPRYVEPIKPGTTLTKKIEYVEVADATAYKE  
FRVELDPSAKIEVYYDKNKKKEETKSFPITEKGFVVPDLSEHIKNPGLITKVI  
EKK"

sig\_peptide  
mat\_peptide

344..415  
416..832

BASE COUNT 365 a 123 c 190 g 345 t  
ORIGIN

alignment\_scores:  
Quality: 695.00 Length: 136  
Ratio: 5.148 Gaps: 0  
Percent Similarity: 99.265 Percent Identity: 97.794

alignment\_block:

US-09-728-670-10 x SASAK42D ..

Align seg 1/1 to: SASAK42D from: 1 to: 1023

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTyr 17  
|||||  
425 TCAAGTTCATTCGACAAAGAAATATAAAAAGCGATGACGCGAGTTA 474  
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
475 TTTTGAACCAACAGCGCCGATTTGATGATAATGTGACTGAGATTGATG 524  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
::|||  
525 GTAAAGAATGTAATGCTATCCCTCGTTATGTCGAGTTTCTATTAA 574  
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67  
|||||  
575 CCTGGAGCTACACTTACAAAAAGAAATTAATACTATGTCGAATGGC 624  
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84  
|||||  
625 ATTAGATCGACAGCATATAAGAGTTTAGAGTAGTTGAATTAGATCA 674  
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100  
|||||  
675 GCGCAAGATCGAAGTCACTTATGATAAGATAAGAAAAAGAAGAA 724  
101 ThrLysSerPheProIleThrGluLysGlyPheValProAspLeu 117  
|||||  
725 ACGAAGTCTTCCCTATACAGAAAAAGGTTTGTGTCCGAGATTATC 774  
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134  
|||||  
775 AGAGCATATTAAAAACCCTGGATTCACCTTAATTACAAGGTTGTATAG 824  
134 luLysLys 136  
|||||  
825 AAAAGAAA 832

seq\_name: gb\_pat:A17537

seq\_documentation\_block:

LOCUS A17537 1023 bp DNA linear PAT 19-APR-1994  
DEFINITION Bacteriophage 42D gene for staphylokinase 42D.  
ACCESSION A17537  
VERSION A17537.1 GI:512341  
KEYWORDS Bacteriophage 42D.  
SOURCE Bacteriophage 42D.  
ORGANISM Bacteriophage 42D  
Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
lambda phage group.  
REFERENCE 1 (bases 1 to 1023)  
AUTHORS  
TITLE EXPRESSION SIGNAL-PEPTIDE-FREE STAPHYLOKINASES  
JOURNAL Patent: WO 9313209-A 9 08-JUL-1993;  
FEATURES  
location/Qualifiers  
1..1023  
/organism="Bacteriophage 42D"  
/db\_xref="taxon:10715"

-35\_signal  
-10\_signal

224..229  
248..253

```
RBS          332. .336
CDS          344. .835
            /codon_start=1
            /transl_table=11
            /product="staphylokinase 42D"
            /protein_id="CA01341.1"
            /db_xref="GI:512342"
            /db_xref="SWISS-PROT:P15240"
            /translation="MLKRSLLFLTVLLLFSSSTITNEVSASSSFDKGYKKGDASY
FEPTGYLMVNTGVGDKRNNELSPRYVEFPKPGTTLTKKLEYVEMALDATAYKE
FRVELDPSAKIEVTYYDKNKKKEETKSEPTTEKGFVVDLSEHIKNPGFNLTIKVVI
EKK"
sig_peptide  344. .415
mat_peptide  416. .832
            /product="staphylokinase 42D"
BASE COUNT   365 a   123 c   190 g   345 t
ORIGIN
alignment_scores:
    quality: 695.00      length: 136
    ratio: 5.148        gaps: 0
    percent similarity: 99.265    percent identity: 97.794
alignment_block:
US-09-728-670-10 x A17537 ..
Align seg 1/1 to: A17537 from: 1 to: 1023
1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17
|||||
425 TCAGTTCATTCGACAAAGAAATATATAAAGCGATGACGCGAGTTA 474
17 rphegluProthrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
475 TTTGAACCAACAGCGCCGATTTGATGTTAAATGTGACTGGAGTTGATG 524
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::|
525 GTAAAGAAATGAATGCTATCCCTCGTTATGTGAGTTTCCCTATTAAA 574
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
|||||
575 CCGGACTTACACTTACAAAAGAAAAATGAAATGACTATGTCGAATGGCC 624
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
625 ATTAGATGCGACGACATATAAGAGTTTAGAGTTGAAATTAGATCCAA 674
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
675 GCGCAAGATCGAAGTCACTTATATGATTAAGATAAGAAAAAGAGAA 724
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuS 117
|||||
725 ACGAAGTCTTCCCTATAACAGAAAAAGTTTGTGTCCAGATTATATC 774
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
775 AGAGCATATTAATAAACCTGATTCAACTTAATTACAAAGGTGTATATAG 824
134 LuLysLys 136
|||||
825 AAAAGAAA 832
seq_name: gb_pat:AR037177
seq_documentation_block:
LOCUS      AR037177      1023 bp      DNA      linear      PAT 29-SEP-1999
DEFINITION Sequence 16 from patent US 5801037.
ACCESSION  AR037177
VERSION     AR037177.1  GI:5955033
KEYWORDS
```

```
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 1023)
AUTHORS     Behnke,D., Schlott,B., Albrecht,S., Guhrs,K.-H. and Hartmann,M.
TITLE       Expression of signal-peptide-free staphylokinases
JOURNAL     Patent: US 5801037-A 16 01-SEP-1998;
FEATURES
    source   location/Qualifiers
            1. .1023
BASE COUNT   365 a   123 c   190 g   345 t
ORIGIN
alignment_scores:
    quality: 695.00      length: 136
    ratio: 5.148        gaps: 0
    percent similarity: 99.265    percent identity: 97.794
alignment_block:
US-09-728-670-10 x AR037177 ..
Align seg 1/1 to: AR037177 from: 1 to: 1023
1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17
|||||
425 TCAGTTCATTCGACAAAGAAATATATAAAGCGATGACGCGAGTTA 474
17 rphegluProthrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
475 TTTGAACCAACAGCGCCGATTTGATGTTAAATGTGACTGGAGTTGATG 524
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::|
525 GTAAAGAAATGAATGCTATCCCTCGTTATGTGAGTTTCCCTATTAAA 574
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
|||||
575 CCGGACTTACACTTACAAAAGAAAAATGAAATGACTATGTCGAATGGCC 624
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84
|||||
625 ATTAGATGCGACGACATATAAGAGTTTAGAGTTGAAATTAGATCCAA 674
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
675 GCGCAAGATCGAAGTCACTTATATGATTAAGATAAGAAAAAGAGAA 724
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuS 117
|||||
725 ACGAAGTCTTCCCTATAACAGAAAAAGTTTGTGTCCAGATTATATC 774
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
775 AGAGCATATTAATAAACCTGATTCAACTTAATTACAAAGGTGTATATAG 824
134 LuLysLys 136
|||||
825 AAAAGAAA 832
seq_name: gb_pat:AR037174
seq_documentation_block:
LOCUS      AR037174      414 bp      DNA      linear      PAT 29-SEP-1999
DEFINITION Sequence 11 from patent US 5801037.
ACCESSION  AR037174
VERSION     AR037174.1  GI:5955030
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 414)
AUTHORS     Behnke,D., Schlott,B., Albrecht,S., Guhrs,K.-H. and Hartmann,M.
```

TITLE Expression of signal-peptide-free staphylokinases  
JOURNAL Patent: US 5801037-A 11 01-SEP-1998;  
FEATURES Location/Qualifiers  
source 1..414  
BASE COUNT 158 a 60 c 81 g 106 t 9 others  
ORIGIN

alignment\_scores:  
Quality: 689.00 Length: 136  
Ratio: 5.142 Gaps: 0  
Percent Similarity: 98.529 Percent Identity: 97.059

alignment\_block:  
US-09-728-670-10 x AR037174 ..

Align seg 1/1 to: AR037174 from: 1 to: 414

```
1 SerSerSerPheAspLysGlySTyrLysLysGlyAspAspAlaSerTy 17
|||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGAGTACGCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTTTGAACCAACAGCGCCGTAATTGTGYTAATGTGACTGAGTYGAYG 103
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
::|||
104 GTAAAGAAATGARCSTYTRTCCCKGKTATGTCGAGTTCTTATTA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
|||||
154 CCTGGGACTACACTTACAAAGAAATAATTGAATACTATGTCGAATGGGC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspProS 84
|||||
204 ATTAGATCGACAGCATATTAAGATTAGAGTAGTGAATTAGATCCAA 253
84 eraLalysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
254 GCGCAAGATCGAAGTCACTTATATGATAAGAATAAGAAAAAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
|||||
304 ACGAAGTCTTCCCTATACAGAAAAAGTTTGTGTGCCAGATTATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
354 AGAGCATATTAATAAACCTCGATTCACTTAATTACCAAGTTGTATAG 403
134 LuLysLys 136
|||||
404 AAAAGAAA 411
```

seq\_name: gb\_pat:AR037175

seq\_documentation\_block:

LOCUS AR037175 414 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 13 from patent US 5801037.  
ACCESSION AR037175  
VERSION AR037175.1 GI:5955031

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 414)

AUTHORS Behnke,D., Schlott,B., Albrecht,S., Guhrs,K.-H. and Hartmann,M.  
TITLE Expression of signal-peptide-free staphylokinases  
JOURNAL Patent: US 5801037-A 13 01-SEP-1998;  
FEATURES location/Qualifiers  
source 1..414  
/organism="unknown"

BASE COUNT 159 a 60 c 80 g 106 t 9 others  
ORIGIN

alignment\_scores:  
Quality: 687.00 Length: 136  
Ratio: 5.127 Gaps: 0  
Percent Similarity: 98.529 Percent Identity: 96.324

alignment\_block:  
US-09-728-670-10 x AR037175 ..

Align seg 1/1 to: AR037175 from: 1 to: 414

```
1 SerSerSerPheAspLysGlySTyrLysLysGlyAspAspAlaSerTy 17
|||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGAGTACGCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTTTGAACCAACAGCGCCGTAATTGTGCTAGTAATGTGACTGAGTYGAYG 103
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
::|||
104 GTAAAGAAATGARCSTYTRTCCCKGKTATGTCGAGTTCTTATTA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
|||||
154 CCTGGGACTACACTTACAAAGAAATAATTGAATACTATGTCGAATGGGC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspProS 84
|||||
204 ATTAGATCGACAGCATATTAAGATTAGAGTAGTGAATTAGATCCAA 253
84 eraLalysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
254 GCGCAAGATCGAAGTCACTTATATGATAAGAATAAGAAAAAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
|||||
304 ACGAAGTCTTCCCTATTAACAGAAAAAGTTTGTGTGCCAGATTATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
354 AGAGCATATTAATAAACCTCGATTCACTTAATTACCAAGTTGTATAG 403
134 LuLysLys 136
|||||
404 AAAAGAAA 411
```

seq\_name: gb\_pat:A17534

seq\_documentation\_block:

LOCUS A17534 414 bp DNA linear PAT 19-APR-1994  
DEFINITION Bacteriophage 42D DNA for staphylokinase SAKM26C.  
ACCESSION A17534  
VERSION A17534.1 GI:512337

KEYWORDS

SOURCE Bacteriophage 42D.  
ORGANISM Bacteriophage 42D  
Viruses: dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;

REFERENCE 1 (bases 1 to 414)

AUTHORS EXPRESSION SIGNAL-PEPTIDE-FREE STAPHYLOKINASES  
JOURNAL Patent: WO 9313209-A 6 08-JUL-1993;  
FEATURES location/Qualifiers  
source 1..414  
/organism="Bacteriophage 42D"  
/db\_xref="taxon:10715"  
/codon\_start=1  
/transl\_table=11

BASE COUNT	158	a	60	c	81	g	106	t	9	others
ORIGIN	/product="staphylokinase SAKM26C" /protein_id="CAA01339.1" /db_xref="GI:512338" /translation="MSSSFDKGKKKGDASYFEPTGPYLXVNVTVGXGRNKLXSPR YVEFPKPGTLLTKEKIEYVEMALDATAKEERVELDPSAKIEVTYYDKNKKKEET KSPITEKGFEVFDLSEHIKNPGENLITKVIEKK"									

```
alignment_scores:
    Quality: 671.00      Length: 136
    Ratio:   5.122       Gaps:     0
Percent Similarity: 96.324 Percent Identity: 94.853
```

```
alignment_block:
US-09-728-670-10 x A17534 ..
```

Align seg 1/1 to: A17534 from: 1 to: 414

1	SetSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTyr	17
4	TCAAGTTCATTCGACCAAGCAAAATATATAAAAAAGCGATGACGCGAGTTA	53
17	rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS	34
54	TTTTTGAAACCAACAGGCCCGTATTTGTGNGTAATATGTGACTGAGTNGANG	103
34	erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys	50
	:::	
104	GTAAGAAGAAATGANCNTNNTNTCCCCNCNGNTATGTGAGTTCCATAATAA	153
51	ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl	67
154	CGTGGCACTACACCTTACAAAAGAAAAATGAATACTATGTGCAATGGGC	203
67	aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS	84
204	ATTAGATGCGACAGCATATATAAGAGCTTAGAGTAGTTGAATTAGATCCAA	253
84	erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu	100
254	GCGCAAGATCGAAGTCACTTATATGATAAGAATAAGAAAAAGAGAA	303
101	ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe	117
304	ACGAAGTCTTCCCTATATACAGAAAAAGTTTGTGTGCCAGATTATC	353
117	rgLuiHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG	134
354	AGAGCATATTAAAAACCCGTGATTCACTTAATTACCAAGGTTGTATTAG	403
134	IuLysLys	136
404	AAAAGAAA	411

seq\_name: gbb\_pat:A17535

seq\_documentation\_block:

LOCUS	A17535	414 bp	DNA	linear	PAT 19-APR-1994
DEFINITION	Bacteriophage 42D DNA for staphylokinase SAKm26L.				
ACCESSION	A17535				
VERSION	A17535.1	GI:512339			

**SOURCE**

ORGANISM	Bacteriophage	42D
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;

REFERENCE 1 (bases 1 to 414)

## REFERENCES

EXPRESSION SIGNAL-PEPTIDE-FREE STAPHYLOKINASES

JOURNAL Patent: WO 9313209-A 7 08-JUL-1993;

**FEATURES**

**Location/Qualifiers**

source	1. .414
--------	---------

**CDS**

BASE COUNT	159	a	60	c	81	g	105	t	9	others	
ORIGIN	/translation="MSSSFDKGKYYYKKKGDDASTFEPTGYLTIVNTGXKGRNXLXSPR YVEEPIKPGTTLTLEKIEIYYVEWALDATALKEFVELDPSAKIEVTVYDKNKKKEET KSPITEKGCVVPDLSEHIKNPGENLITKVVIEKK"										

[illegible]

```
alignment_block:
US-09-728-670-10 x A17535 .
```

Align seg 1/1 to: A17535 from: 1 to: 414

1	SetSerSerPheaspIysGlyIysTyrIysLysGlyaspaspAlaSerTy	17
4	TCAAGTTCATTCGACCAAGGAAATATATAAAAGGCGATGACCGCGACTTA	53
17	rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsps	34
54	TTTTGAACCAACACAGGCCCGTATTTGACGCGTAATGTGACTGAGNTGANG	103
34	erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys	50
	::	
104	GTAAGAGAAATGANCTNNTNTCCCCNCNGNTATGTGCGATTTCCTATTAA	153
51	ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl	67
154	CCTGGACTACACTTACAAAAGAAAAATTGAATACTATGTGCAATGGCC	203
67	aleuAspAlaThrAlaTyrLysGluPheargValGluLeuAspProS	84
204	ATTAGATCGACAGCATATAAAGAGTTTAGAGTAGTTGAATTAGATCCAA	253
84	erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu	100
254	GCGCAAGATCGAAGTCACCTTATATGATAAGATAAGAAAAAGAGAA	303
101	ThrLysSerPheProIleThrGluLysGlyPheValValProaspLeuSe	117
304	ACGAAGTCCTTCCCTATAACAGAAAAAGTTTGTGTGCCACAGATTATC	353
117	rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG	134
354	AGAGCATATTAAAAACCCCTGGATTCACTTAATTACNAAGTTGTTATAG	403
134	IuLysLys	136
404	AAAAGAA	411

seq\_name: gjb\_pat:A17532

seq\_documentation\_block:

LOCUS	A17532	384 bp	DNA	linear	PAT 06-JUL-1995
-------	--------	--------	-----	--------	-----------------

**DEFINITION** Bacteriophage 42D DNA for staphylokinase SAK-delta-N10 (short).

```

RECESSION  817002
VERSION    A17532.1  GI:512333

```

KEYWORD

SOURCE Bacteriophage 42D.

ORGANISM Bacteriophage 42D

Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;



REFERENCE 1 (bases 1 to 384)  
LAMBDA PHAGE GROUP.  
AUTHORS  
TITLE  
JOURNAL  
PATENT: WO 9313209-A 4 08-JUL-1993;  
FEATURES  
LOCATION/QUALIFIERS  
1. 384  
/ORGANISM="BACTERIOPHAGE 42D"  
/DB\_XREF="TAXON:10715"  
/CODON\_START=1  
/NOTE="SHORT FORM"  
/TRANSI\_TABLE=11  
/PRODUCT="STAPHYLOKINASE SAK-DELTA-N10"  
/PROTEIN\_ID="CAA01337.1"  
/DB\_XREF="GI:512334"  
/TRANSLATION="MKGDASYFEPTGPYLMVNTGVGDKRNLSPRYVEPIKPGT  
TLTKREIYVEMALDATAYKEFRVELDPSAKIEVTYYDKNKKKEETKSPITEKGF  
VVPDLSEHITKNPGFNLTITKVIEK"

BASE COUNT 146 A 55 C 78 G 104 T 1 OTHERS  
ORIGIN

ALIGNMENT\_SCORES:  
QUALITY: 643.00 LENGTH: 126  
RATIO: 5.144 GAPS: 0  
PERCENT SIMILARITY: 99.206 PERCENT IDENTITY: 97.619

ALIGNMENT\_BLOCK:  
US-09-728-670-10 X A17532 ..

ALIGN SEG 1/1 TO: A17532 FROM: 1 TO: 384

11 LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVa 27  
|||||  
4 AAAGCGATGACGCGAGTTATTTGAACCAACAGAGCCCGTATTGATGGT 53  
27 IAsnValThrGlyValAspSerLysGlyAsnGluLeuSerProHisT 44  
|||||  
54 AAATGTGACTGGAGTTGATGTAAGAAATGAATGCTATCCCTCGTT 103  
44 YrValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIle 60  
|||||  
104 ATGTCGAGTTTCCCTATTAACTGGGACTACACTTACAAAGAAATAAT 153  
61 GluTyrTyrValGluTrrPalaLeuAspAlaThrAlaTyrLysGluPheAr 77  
|||||  
154 GAATTAAGTATGTCGATGGGCAATTAGATCGACAGCATTAAGAAGTTAG 203  
77 GValValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspL 94  
|||||  
204 AGTAGTTGAATTAGATCCAAAGCGCAAGATCGAAGTCACTTATTATGATA 253  
94 YsAsnLysLysLysGluGluThrLysSerPheProIleThrGluLysGly 110  
|||||  
254 AGAATTAAGAAAAAGAAAGAAAGAGTCTTCCCTATTAAACAGAAAAAGGT 303  
111 PheValValProAspLeuSerGluHisIleLysAsnProGlyPheAsnLe 127  
|||||  
304 TTTGTGTCCAGATTATTCAGACATATTAAAAACCCCTGAGTTCAACTT 353  
127 uIleThrLysValValIleGluLysLys 136  
|||||  
354 AATTACNAAGGTGTTATAGAAAAAGAAA 381

seq\_name: gb\_pat:AR037172

seq\_documentation\_block:  
LOCUS AR037172 384 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 7 from patent US 5801037.  
ACCESSION AR037172  
VERSION AR037172.1 GI:5955028  
KEYWORDS

SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 384)  
AUTHORS  
TITLE  
JOURNAL  
PATENT: US 5801037-A 7 01-SEP-1998;  
FEATURES  
LOCATION/QUALIFIERS  
1. 384  
/ORGANISM="UNKNOWN"

BASE COUNT 146 A 55 C 78 G 104 T 1 OTHERS  
ORIGIN

ALIGNMENT\_SCORES:  
QUALITY: 643.00 LENGTH: 126  
RATIO: 5.144 GAPS: 0  
PERCENT SIMILARITY: 99.206 PERCENT IDENTITY: 97.619

ALIGNMENT\_BLOCK:  
US-09-728-670-10 X AR037172 ..

ALIGN SEG 1/1 TO: AR037172 FROM: 1 TO: 384

11 LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVa 27  
|||||  
4 AAAGCGATGACGCGAGTTATTTGAACCAACAGAGCCCGTATTGATGGT 53  
27 IAsnValThrGlyValAspSerLysGlyAsnGluLeuSerProHisT 44  
|||||  
54 AAATGTGACTGGAGTTGATGTAAGAAATGAATGCTATCCCTCGTT 103  
44 YrValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIle 60  
|||||  
104 ATGTCGAGTTTCCCTATTAACTGGGACTACACTTACAAAGAAATAAT 153  
61 GluTyrTyrValGluTrrPalaLeuAspAlaThrAlaTyrLysGluPheAr 77  
|||||  
154 GAATTAAGTATGTCGATGGGCAATTAGATCGACAGCATTAAGAAGTTAG 203  
77 GValValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspL 94  
|||||  
204 AGTAGTTGAATTAGATCCAAAGCGCAAGATCGAAGTCACTTATTATGATA 253  
94 YsAsnLysLysLysGluGluThrLysSerPheProIleThrGluLysGly 110  
|||||  
254 AGAATTAAGAAAAAGAAAGAAAGAGTCTTCCCTATTAAACAGAAAAAGGT 303  
111 PheValValProAspLeuSerGluHisIleLysAsnProGlyPheAsnLe 127  
|||||  
304 TTTGTGTCCAGATTATTCAGACATATTAAAAACCCCTGAGTTCAACTT 353  
127 uIleThrLysValValIleGluLysLys 136  
|||||  
354 AATTACMAAGGTGTTATAGAAAAAGAAA 381

seq\_name: gb\_pat:A17533

seq\_documentation\_block:  
LOCUS A17533 372 bp DNA linear PAT 19-APR-1994  
DEFINITION Bacteriophage 42D DNA for staphylokinase SAK-delta-N14 (short).  
ACCESSION A17533  
VERSION A17533.1 GI:512335  
KEYWORDS  
SOURCE Bacteriophage 42D.  
ORGANISM Bacteriophage 42D  
VIRUSES; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
LAMBDA PHAGE GROUP.

REFERENCE 1 (bases 1 to 372)  
AUTHORS  
TITLE  
JOURNAL  
PATENT: WO 9313209-A 5 08-JUL-1993;  
FEATURES  
LOCATION/QUALIFIERS

```
source          1..372
                /organism="Bacteriophage 42D"
                /db_xref="taxon:10715"
CDS
  1..372
    /note="short form"
    /codon_start=1
    /transl_table=11
    /product="staphylokinase SAK-delta-N14"
    /protein_id="CAA01338.1"
    /db_xref="GI:512336"
    /translation="MASYFEPTGPLYLMVNTGVDGKRNELSPRYVEPIKPGTTLTK
EKIEYVEMALDATAVKEFRVELDPSAKIEVTYYDKNKKKEETKSPFITEKGFVVPD
LSEHIKNPGFNLTFRKVIIEKK"
BASE COUNT      141 a      53 c      74 g     103 t      1 others
ORIGIN

alignment_scores:
  Quality: 620.00      Length: 122
  Ratio: 5.124        Gaps: 0
  Percent Similarity: 99.180      Percent Identity: 97.541

alignment_block:
US-09-728-670-10 x A17533      ..
Align seg 1/1  to: A17533  from: 1  to: 372

15  AlasertyrphegluProthrGlyProtyrleuMetValasnValThrGl 31
|||||
4  GCGAGTTATTGTAACCAACAGAGCCCGTATTGATGCTAAATGTGACTGG 53

31  yValasPserLysGlyAsnGluLeuLeuSerProHisTyrValGluPhe 48
|||||
54  AGTGATGCTAAAGAAATGAATTGCTATCCCTCGTTATGTCGAGTTTC 103

48  roileLysProGlyThrThrLeuThrLysGluLysileGluTyrTyrVal 64
|||||
104 CTATTAAACCTGGGACTACACTTACAAAGAAAAAATTGAATACTATGTC 153

65  GluTrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValGluLe 81
|||||
154 GAATGGGCATTAGATGCGACAGCATATTAAGAGTTTACAGTACTTGAATP 203

81  uAspProSerAlaLysileGluValThrTyrTyrAspLysAsnLysLysL 98
|||||
204 AGATCCAAAGCGCAAGATCGAAGTCACTTATGATTAAGAATAAGAAAA 253

98  ysgLugluThrLysSerPheProileThrGluLysGlyPheValValPro 114
|||||
254 AAGAAGAAACGAAGCTTTCCCTATTAACAGAAAAAGTTTGTGTGCCCA 303

115  AspleuSerGluHisileLysAsnProGlyPheAsnLeuileThrLysVa 131
|||||
304 GATTATCAGAGCATATTAAAAACCCGTGATTCACACTTAATTACNAAGCT 353

131  lValileGluLysLys 136
|||||
354 TGTATTAGAAAAAGAAA 369

seq_name: gb_pat:AR037173

seq_documentation_block:
LOCUS      AR037173      372 bp      DNA      linear      PAT 29-SEP-1999
DEFINITION Sequence 9 from patent US 5801037.
ACCESSION  AR037173
VERSION    AR037173.1  GI:5955029
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 372)
AUTHORS   Behnke,D., Schlott,B., Albrecht,S., Guhrs,K.-H. and Hartmann,M.
TITLE     Expression of signal-peptide-free staphylokinases
```

```
JOURNAL      Patent: US 5801037-A 9 01-SEP-1998;
FEATURES
  source      1..372
              /organism="unknown"
BASE COUNT    141 a      53 c      74 g     103 t      1 others
ORIGIN

alignment_scores:
  Quality: 620.00      Length: 122
  Ratio: 5.124        Gaps: 0
  Percent Similarity: 99.180      Percent Identity: 97.541

alignment_block:
US-09-728-670-10 x AR037173      ..
Align seg 1/1  to: AR037173  from: 1  to: 372

15  AlasertyrphegluProthrGlyProtyrleuMetValasnValThrGl 31
|||||
4  GCGAGTTATTGTAACCAACAGAGCCCGTATTGATGCTAAATGTGACTGG 53

31  yValasPserLysGlyAsnGluLeuLeuSerProHisTyrValGluPhe 48
|||||
54  AGTGATGCTAAAGAAATGAATTGCTATCCCTCGTTATGTCGAGTTTC 103

48  roileLysProGlyThrThrLeuThrLysGluLysileGluTyrTyrVal 64
|||||
104 CTATTAAACCTGGGACTACACTTACAAAGAAAAAATTGAATACTATGTC 153

65  GluTrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValGluLe 81
|||||
154 GAATGGGCATTAGATGCGACAGCATATTAAGAGTTTACAGTACTTGAATP 203

81  uAspProSerAlaLysileGluValThrTyrTyrAspLysAsnLysLysL 98
|||||
204 AGATCCAAAGCGCAAGATCGAAGTCACTTATGATTAAGAATAAGAAAA 253

98  ysgLugluThrLysSerPheProileThrGluLysGlyPheValValPro 114
|||||
254 AAGAAGAAACGAAGCTTTCCCTATTAACAGAAAAAGTTTGTGTGCCCA 303

115  AspleuSerGluHisileLysAsnProGlyPheAsnLeuileThrLysVa 131
|||||
304 GATTATCAGAGCATATTAAAAACCCGTGATTCACACTTAATTACMAAGCT 353

131  lValileGluLysLys 136
|||||
354 TGTATTAGAAAAAGAAA 369

seq_name: gb_pat:E00876

seq_documentation_block:
LOCUS      E00876      634 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION DNA fragment comprising a promoter of sak gene and the region
              coding the signal peptide.
ACCESSION  E00876
VERSION    E00876.1  GI:2169137
KEYWORDS   JP 1986135590-A/1.
SOURCE     Staphylococcus aureus.
ORGANISM   Staphylococcus aureus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE  1 (bases 1 to 634)
AUTHORS   Sako,T.
TITLE     COMPOSITE PLASMID
JOURNAL    Patent: JP 1986135590-A 1 23-JUN-1986;
            YAKULT HONSHA CO LTD
COMMENT    OS Staphylococcus aureus
            PN JP 1986135590-A/1
            PD 23-JUN-1986
            PF 05-DEC-1984 JP 1984257258
            PI SAKO TOMOYUKI
```

PC C12N15/00//C07H21/04,(C12N15/00,C12R1:445);  
CC strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
FH key Location/Qualifiers  
FT 3'UTR <1..300  
FT promoter 189..220  
FT RBS 299..303  
FT sig\_peptide 311..391  
FT variation replace(405..406,'CG',407..634,'') FT  
variation replace(392,'G',393..634,'') FT variation  
replace(405,'G',406..634,'').  
Location/Qualifiers  
1..634  
/organism="Staphylococcus aureus"  
/db\_xref="taxon:1280"  
BASE COUNT 221 a 67 c 117 g 229 t  
ORIGIN

alignment\_scores:  
Quality: 426.00 Length: 81  
Ratio: 5.259 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.765

alignment\_block:  
US-09-728-670-10 x E00876 ..  
Align seg 1/1 to: E00876 from: 1 to: 634

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTy 17  
|||||  
392 TCAAGTTCATTCGACAAAGAAATATATAAAAGGGGATGACGCGAGTTA 441  
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
442 TTTTGAACCAACAGCGCCCGTATTGATGTAAATGTGACTGGAGTTGATG 491  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
:::|||||  
492 GTAAAGAAATGAATGCTATCCCTCATTTATGTCGAGTTTCTATTAAA 541  
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrAl 67  
|||||  
542 CCTGGAGCTACACTTACAAAGAAAAAATGAATACTATGTCGAATGGGC 591  
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeu 81  
|||||  
592 ATTAGATGCGACAGCATATAAGAGTTTAGAGTAGTTGAATTA 634

seq\_name: gb\_pat:E02874

seq\_documentation\_block:  
LOCUS E02874 480 bp DNA linear PAT 29-SEP-1997  
DEFINITION DNA encoding fusion somatomedin C.  
ACCESSION E02874  
VERSION E02874.1 GI:2171099  
KEYWORDS JP 1991098595-A/3.  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequence.  
1 (bases 1 to 480)  
REFERENCE Hanada,K., Aoyama,J., Ikeda,Y., Yoshimoto,M., Yoshimura,H.,  
AUTHORS Tamaki,S. and Kondo,I.  
TITLE PRODUCTION OF PEPTIDES IN GENE ENGINEERING  
JOURNAL Patent: JP 1991098595-A 3 24-APR-1991;  
TAISHO PHARMACEUT CO LTD  
OS Artificial gene  
OC Artificial sequence; Genes.  
PN JP 1991098595-A/3

PD 24-APR-1991  
PF 11-SEP-1989 JP 1989234874  
PI HANADA KAZUNORI, AOYAMA JUNKO, IKEDA YOKO, YOSHIMOTO MAKOTO,  
PI YOSHIMURA HIROMITSU, TAMAKI SHIGEO, KONDO ISAMU PC  
C12P21/00,C12N1/21,C12N15/62,C12N15/70,C12N15/75,(C12P21/00, PC  
C12R1:19),  
PC (C12P21/00,C12R1:125);  
CC strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
CC \*source: clone=PKS002;  
FH key Location/Qualifiers  
FT CDS 1..477  
FT sig\_peptide 1..81  
FT /product='precursor fusion somatomedin C' FT  
FT mat\_peptide 82..474  
FT /note='SAK signal peptide'  
FT /note='fusion somatomedin C'  
FT misc\_feature 1..236  
FT /note='N-terminal of precursor FT  
FT staphylokinase' 237..264  
FT misc\_feature /note='BamHI linker'  
FT misc\_feature 265..477  
FT /note='somatomedin C'  
FT 3'UTR 478..480.  
Location/Qualifiers  
1..480  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 125 a 101 c 108 g 146 t  
ORIGIN

alignment\_scores:  
Quality: 277.00 Length: 52  
Ratio: 5.327 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.077

alignment\_block:  
US-09-728-670-10 x E02874 ...  
Align seg 1/1 to: E02874 from: 1 to: 480

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17  
|||||  
82 TCAAGTTCATTCGACAAAGAAATATAAAAGGGGATGACGCGAGTTA 131  
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
132 TTTTGAACCAACAGCGCCCGTATTGATGTAAATGTGACTGGAGTTGATG 181  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
:::|||||  
182 GTAAAGAAATGAATGCTATCCCTCATTTATGTCGAGTTTCTATTAAA 231  
51 ProGly 52  
|||||  
232 CCTGGG 237

seq\_name: gb\_pat:E03544

seq\_documentation\_block:  
LOCUS E03544 492 bp DNA linear PAT 29-SEP-1997  
DEFINITION DNA sequence coding for staphylokinase signal peptide,  
staphylokinase N-terminus peptide and tyrostatin.  
ACCESSION E03544  
VERSION E03544.1 GI:2171760  
KEYWORDS JP 1992112791-A/3.  
SOURCE synthetic construct.  
ORGANISM synthetic construct

artificial sequence.  
REFERENCE 1 (bases 1 to 492)  
AUTHORS Yoshimoto, M., Hanada, K. and Tamaki, S.  
TITLE TRYPSTATIN STRUCTURAL GENE AND USE THEREOF  
JOURNAL Patent: JP 1992112791-A 3 14-APR-1992;  
TAISHO PHARMACEUT CO LTD  
COMMENT OS Artificial gene  
OC Artificial sequence; Genes.  
PN JP 1992112791-A/3  
PD 14-APR-1992  
PF 31-AUG-1990 JP 1990232268  
PI YOSHIMOTO MAKOTO, HANADA KAZUNORI, TAMAKI SHIGEO PC  
C12N15/15, C12N1/21, C12N9/99, C12N15/62, (C12N1/21, C12R1:19), PC  
(C12N1/21,  
PC C12R1:865), (C12N9/99, C12R1:19), (C12N9/99, C12R1:865); CC  
strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
FH Key Location/Qualifiers  
FT 5'UTR 1..51  
FT misc\_signal 40..44  
FT /note='shine Dalgarno sequence' FT  
sig\_peptide 52..132 /product='staphylokinase signal peptide' FT  
FT 133..300  
CDS /product='staphylokinase'  
FT 301..486  
FT /product='trypstatin'  
FT 487..492.  
FT 3'UTR Location/Qualifiers  
FEATURES  
source 1..492  
/organism='synthetic construct'  
/db\_xref='taxon:32630'  
BASE COUNT 151 a 90 c 107 g 144 t  
ORIGIN  
alignment\_scores:  
Quality: 277.00 Length: 52  
Ratio: 5.327 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.077  
alignment\_block:  
US-09-728-670-10 x E03544 ..  
Align seg 1/1 to: E03544 from: 1 to: 492  
1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17  
|||||  
133 TCAAGTTCATTCGACAAAGAAATATATAAAGCGATGACGCGAGTTA 182  
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
183 TTTGAACCAACAGAGCGCGTATTGTGATGTAATGTGACTGAGTTGATG 232  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
:::|||||  
233 GTAAAGGAATGAATTGCTATCCCTCATTTATGTGAGTTCTCTATTAA 282  
51 ProGly 52  
|||||  
283 CCTGGG 288  
seq\_name: gb\_pat:E03543  
seq\_documentation\_block:  
LOCUS E03543 494 bp DNA linear PAT 29-SEP-1997  
DEFINITION DNA sequence coding for staphylokinase signal peptide,  
staphylokinase N-terminus peptide and trypstatin.  
ACCESSION E03543  
VERSION E03543.1 GI:2171759

JP 1992112791-A/2.  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 494)  
AUTHORS Yoshimoto, M., Hanada, K. and Tamaki, S.  
TITLE TRYPSTATIN STRUCTURAL GENE AND USE THEREOF  
JOURNAL Patent: JP 1992112791-A 2 14-APR-1992;  
TAISHO PHARMACEUT CO LTD  
COMMENT OS Artificial gene  
OC Artificial sequence; Genes.  
PN JP 1992112791-A/2  
PD 14-APR-1992  
PF 31-AUG-1990 JP 1990232268  
PI YOSHIMOTO MAKOTO, HANADA KAZUNORI, TAMAKI SHIGEO PC  
C12N15/15, C12N1/21, C12N9/99, C12N15/62, (C12N1/21, C12R1:19), PC  
(C12N1/21,  
PC C12R1:865), (C12N9/99, C12R1:19), (C12N9/99, C12R1:865); CC  
strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
FH Key Location/Qualifiers  
FT 5'UTR 1..59  
FT misc\_signal 48..52  
FT /note='shine Dalgarno sequence' FT  
sig\_peptide 60..140  
FT 141..299  
mat\_peptide /product='staphylokinase signal peptide' FT  
FT 300..488  
FT /product='trypstatin'  
FT 489..494.  
FT 3'UTR Location/Qualifiers  
FEATURES  
source 1..494  
/organism='synthetic construct'  
/db\_xref='taxon:32630'  
BASE COUNT 152 a 91 c 106 g 145 t  
ORIGIN  
alignment\_scores:  
Quality: 277.00 Length: 52  
Ratio: 5.327 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.077  
alignment\_block:  
US-09-728-670-10 x E03543 ..  
Align seg 1/1 to: E03543 from: 1 to: 494  
1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17  
|||||  
141 TCAAGTTCATTCGACAAAGAAATATATAAAGCGATGACGCGAGTTA 190  
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
191 TTTGAACCAACAGAGCGCGTATTGTGATGTAATGTGACTGAGTTGATG 240  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
:::|||||  
241 GTAAAGGAATGAATTGCTATCCCTCATTTATGTGAGTTCTCTATTAA 290  
51 ProGly 52  
|||||  
291 CCTGGG 296  
seq\_name: gb\_da:SUB314852  
seq\_documentation\_block:  
LOCUS SUB314852 2058 bp DNA linear BCT 21-DEC-2001  
DEFINITION Streptococcus uberis paub gene, ORF2 and ORF3.



ACCESSION AJ314852  
VERSION AJ314852.1 GI:16215456  
KEYWORDS ORF2; ORF3; paub gene; plasminogen activator.  
SOURCE Streptococcus uberis.  
ORGANISM Streptococcus uberis  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (sites)  
AUTHORS Ward, P.N. and Leigh, J.A.  
TITLE Characterization of paub, a Novel Broad-Spectrum Plasminogen  
JOURNAL Activator from Streptococcus uberis  
PUBMED J. Bacteriol. 184 (1), 119-125 (2002)  
11741851  
REFERENCE 2 (bases 1 to 2058)  
AUTHORS Ward, P.N.  
TITLE Direct Submission  
JOURNAL Submitted (08-JUN-2001) Ward P.N., Environmental Microbiology,  
Institute for Animal Health, Compton Laboratory, Compton, nr  
Newbury, Berkshire, RG20 7NN, UNITED KINGDOM  
location/Qualifiers  
FEATURES  
source 1. 2058  
/organism="Streptococcus uberis"  
/strain="SK880"  
/db\_xref="taxon:1349"  
/country="Denmark"  
109. 1242  
/gene="paub"  
109. 183  
/gene="paub"  
109. 1242  
/gene="paub"  
/function="bovine plasminogen activation"  
/codon\_start=1  
/transl\_table=11  
/product="plasminogen activator"  
/protein\_id="CAC85651.1"  
/db\_xref="GI:16215457"  
/translation="MKLFTVLIFIGMGLGIPYKVEATTSKEVNYOEPQMTLTIN  
ITGESKEGOLLSPYEMQFRLTAGEVINKNELLDKVQLVIDSASNOFEVDFKPEK  
VEMSYFDRHLELDITERGEFIVPDYSIEKKPSFLITGPVLIORSPTEKKTIDSNTD  
VILRHVSFMRKNGESMNLVPTTQLSFSKTAKPGQSIQSEELYQAQSLFEQTKY  
HEGYOLIKRLNTSISENEKAYRSVYQNGKPNYTIISOLREFSKWDYSQAIRDDIIE  
NYISKNGDDYPEQIQITLDTNEETSIRTLTLDLPKENTTEHIKHYLSEADLTEI  
DAKTGDHYWFSQNVKQVSSHFFQANYPFSKENSDDLFFYNRFSLQ"  
184. 1239  
mat\_peptide  
/gene="paub"  
/evidence="experimental"  
/product="plasminogen activator"  
complement(1295. 1519)  
/function="unknown"  
/note="ORF2"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAC85652.1"  
/db\_xref="GI:16215458"  
/translation="MLVIGLAPFLLSFLTLLYCLGILSPFYIFLAHVHAGGICIGD  
FYVYLLVFKFKOKRILIEDTLGSLIYQK"  
complement(1556. 1852)  
/function="unknown"  
/note="ORF3"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAC85653.1"  
/db\_xref="GI:16215459"  
/translation="MKIIDKVRLOGNKKLLLLNLSPLATYLFLLFTQIAIIFKPL  
INKGTSINTSGIVTMGMVMAALVIELVHGLFYKLFKPKNRKIKIGFRKSMIG"  
BASE COUNT 762 a 362 c 319 g 615 t  
ORIGIN  
alignment\_scores:

Quality: 231.50 Length: 161  
Ratio: 2.338 Gaps: 2  
Percent Similarity: 61.491 Percent Identity: 31.056  
alignment\_block:  
US-09-728-670-10 x SUB314852 ..  
Align seg 1/1 to: SUB314852 from: 1 to: 2058  
1 SerSerSerPheAspLysGlyLysTyrLysLys..... 11  
1132 CATCTTATTGGAATGTTGGACTACTGGCAATTCCATACAAAGTAGAG 181  
12 .....GlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeu 25  
182 CTATCAGTTCAAAAGAAGTAAATATACCAAGAACACAGGAGGACTATGTTA 231  
26 MetValAsnValThrGlyValAspSerLysGlyAsnGluLeuSerPr 42  
232 ACGATAATATATACCGAGAAAGCAAGACGACAGTGTCTCTTCTCC 281  
42 OHISTYrValGluPheProIleLysProGlyThrThrLeuThrLysGluL 59  
282 AGAATATATGCAATTCGCTTACACGCGGAGAAAGTTATTAATAAATG 331  
59 ysileGluTyrTyrValGluTrpAlaLeuAspAlaThrAlaTyrLysGlu 75  
332 AGTTACTGATTAAGTACCACTGTTATPAGACAGTGCTGCTTCAACAG 381  
76 PheArgValValGluLeuAspProSerAlaLysileGluValThrTyrTy 92  
382 TTTGAAGTTGTTGATTTCAAAACCGCAATCAAAAGGTGAATGCTCATTTT 431  
92 rAspLysAsnLysLysLysGluGluThrLysSerPheProIleThrGluL 109  
432 TGACCGCCACGAAATTGCTTGAT.....ATCACTGAAA 463  
109 ysgLypheValValProAspLeuSerGluHisileLysAsnProGlyPhe 125  
464 GAGGATTTATCGTACTGATTATTCAAATTATGAAAGAAACCTTCATT 513  
126 AsnLeuIleThrLysValValileGluLysLys 136  
514 CTCTTAACAGGCCCTGTTATTATTACAGAGAAA 546  
seq\_name: gb\_ba:AF104301  
seq\_documentation\_block:  
LOCUS AF104301 1248 bp DNA linear BCT 22-NOV-1999  
DEFINITION Streptococcus equisimilis strain 87-542-W streptokinase precursor  
(skc) gene, partial cds.  
ACCESSION AF104301  
VERSION AF104301.1 GI:3983432  
KEYWORDS  
SOURCE  
ORGANISM Streptococcus dysgalactiae subsp. equisimilis.  
Streptococcus dysgalactiae subsp. equisimilis  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 1248)  
AUTHORS Caballero, A.R., Lottenberg, R. and Johnston, K.H.  
TITLE Cloning, expression, sequence analysis, and characterization of  
streptokinases secreted by porcine and equine isolates of  
Streptococcus equisimilis  
JOURNAL Infect. Immun. 67 (12), 6478-6486 (1999)  
MEDLINE 20038313  
REFERENCE 2 (bases 1 to 1248)  
AUTHORS Caballero, A.R., Lottenberg, R. and Johnston, K.H.  
TITLE Direct Submission  
JOURNAL Submitted (04-NOV-1998) MIP, LSU Medical Center, 1901 Perdido



alignment\_block:  
US-09-728-670-10 x AX073720 ..

Align seg 1/1 to: AX073720 from: 1 to: 69

83 ProSerAlaLysIleGluValThrTyrTrpAspLysAsnLysLysGlu 99  
|||||  
1 CCAAGCGCTAAGATCGAAGTCACTATTATGAT.....ATAAAGA 41

99 uGluThrLysSerPheProIleThrGluLys 109  
|||||  
42 AGAAGCG...TCTTCCCTATTAACAGAAAAA 69

seq\_name: gb\_pl:ZMU90944

seq\_documentation\_block:

LOCUS ZMU90944 2004 bp mRNA linear PLN 13-JAN-1999

DEFINITION Zea mays PDI-like protein mRNA, complete cds.

ACCESSION U90944

VERSION U90944.1 GI:4151124

KEYWORDS

SOURCE

ORGANISM Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoidae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 2004)

AUTHORS Laughner,B.J., Sehake,P.C. and Ferl,R.J.

TITLE A novel nuclear member of the thioredoxin superfamily

JOURNAL Plant Physiol. 118 (3), 987-996 (1998)

MEDLINE 99026300

REFERENCE 2 (bases 1 to 2004)

AUTHORS Laughner,B. and Ferl,R.

TITLE Direct Submission

JOURNAL Submitted (26-FEB-1997) Horticultural Sciences, University of

Florida, 1301 Fifield Hall, Gainesville, FL 32611, USA

REFERENCE 3 (bases 1 to 2004)

AUTHORS Laughner,B. and Ferl,R.

TITLE Direct Submission

JOURNAL Submitted (12-JAN-1999) Horticultural Sciences, University of

Florida, 1301 Fifield Hall, Gainesville, FL 32611, USA

REMARK COMMENT On Jan 13, 1999 this sequence version replaced gi:4056567.

FEATURES

source

CDS

1..2004

/organism="Zea mays"

/db\_xref="taxon:4577"

/cell\_type="P3377 suspension cell"

119..1828

/codon\_start=1

/product="PDI-like protein"

/protein\_id="AAD04231.1"

/db\_xref="GI:4056568"

/translation="MAEAPAAAGCGIGDILATSDRDLVRNSGEQYKISSVEASPAI

YFSASWCPGCRFRTPKLIIEVEKLGKSEFVFASADRNENEFYFAKMPWLAV

FSDSEGREALDQFKVSGIPHLVILDAKTGEVTEDEVSEVGEVAYPTPRINP

LKEQEKAEKENTIQSVLVTSTRDYLSNKGDKVPISLEGKYGLCFVVDGFPPLAE

FTEVLAKIYEKLEKEVEKEFVAVSLDSESSFNESFAKMPWLAIPOGDIKCQTLVRY

FELSLPTLVILGPDGKTLNNVADIDHGFEMEGFPFSAEKLIELAEKAKIKAA

QTLSEILISGHLDFVIGKDGAKVPVSELVGTVLVYFSGKWCPCRAFLPTLVKEYNK

IKENSDSEIVFISDEDOSSFDFFSEMPWLAVPWEDERKASLKTFFKIRGIPSLVA

IGPTGQTVSRDAKSQLMIHGADAFPTFERLELQKKLDEMAKGNPQKLKHLHDEHE

LVLRRGTYRDCGNCNEMSSWSYRDCEDFDLHPKCALGEKMGEEAPAGYVCEGVC

RKA"

BASE COUNT 500 a 450 c 544 g 510 t

ORIGIN

alignment\_scores: Quality: 88.00 Length: 109  
Ratio: 1.492 Gaps: 5  
Percent Similarity: 54.128 Percent Identity: 28.440

alignment\_block:  
US-09-728-670-10 x ZMU90944 ..

Align seg 1/1 to: ZMU90944 from: 1 to: 2004

23 ProTyrLeuMetValAsnValThrGlyValAspSerLysGlyAsnGlu 39  
|||||  
407 CCATGGTGGCTGTCCTCTTCT.....GACAGTGAAGGCGGTGAGGC 450

39 uLeuSer.....ProHisTyrValGluPhe 48  
|||||  
451 CCTTGATGGCCAGTTCAGAGTCTCTGTATCCACACACCTGCTCCTGG 500

48 rolleLysProGlyThrThrLeuThrLysGluLysIleGluTyrTyrVal 64  
|||||  
501 ATGCAAAAACCTGGTGAAGTTTACACTGAGATGGTGTGAATTCGTGAGT 550

65 GluTrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValValGlu 81  
|||||  
551 GAGTATGGCGGTGAGGCTTACCTTTACACACGATCGGATCATGAAGT 600

81 uAspProSerAlaLysIleGluValThrTyrTrpAspLysAsnLysLysL 98  
|||||  
601 GAAGAACACAGAAAGGACAGAA.....AAGAGAAATCAAACTA 638

98 ysGlu.....GluThrLysSerPheProIleThrGluLys 109  
|||||  
639 TTCAAAGCGTGTGTGCATCATCTGAGACTATCTCATTTCAACAAG 688

110 GlyPheValValProAspLeuSerGlu 118  
|||||  
689 GGAGACAAAGGTACCC...ATCTCTGAG 712

seq\_name: gb\_pat:AX073717

seq\_documentation\_block:

LOCUS AX073717 66 bp DNA linear PAT 06-FEB-2001

DEFINITION Sequence 10 from Patent WO0104287.

ACCESSION AX073717

VERSION AX073717.1 GI:12710135

KEYWORDS

SOURCE

ORGANISM synthetic construct.  
artificial sequence.

REFERENCE 1 (bases 1 to 66)

AUTHORS Halkier,T., Pedersen,A.H. and Okkels,J.S.

TITLE A method for preparing modified polypeptides

JOURNAL Patent: WO 0104287-A 10 18-JAN-2001;

Maxygen Aps (DK)

FEATURES

source

1..66

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="primer"

BASE COUNT 18 a 11 c 8 g 29 t

ORIGIN

alignment\_scores: Quality: 87.00 Length: 18

Ratio: 4.833 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 94.444

alignment\_block:

US-09-728-670-10 x AX073717/rev ..

Align seg 1/1 to reverse of: AX073717 from: 1 to: 66

119 HisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleGlu 135  
|||||  
65 CATATTAAAAACCTGATTCAGTTAATTACAAAGGTTGTATAGAAAA 16

135 slys 136  
1111  
15 GAAA 12

seq\_name: gb\_ba:AB007638

seq\_documentation\_block:

LOCUS AB007638 16585 bp DNA linear BCT 13-FEB-1999  
DEFINITION Bacillus subtilis genomic DNA containing guta to cota region, 48  
degree.

ACCESSION AB007638  
VERSION AB007638.1 GI:2522006

KEYWORDS  
SOURCE Bacillus subtilis (strain:Marburg 168) DNA.  
ORGANISM Bacillus subtilis  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Bacillus.

REFERENCE 1 (sites)  
AUTHORS Kasahara,Y., Nakai,S., Ogasawara,N., Yata,K. and Sadale,Y.  
TITLE Sequence analysis of the groESL-cota region of the Bacillus  
subtilis genome, containing the restriction/modification system  
genes

JOURNAL DNA Res. (1997) In press  
REFERENCE 2 (bases 1 to 16585)  
AUTHORS Sadale,Y.  
TITLE Direct Submission

JOURNAL Submitted (03-OCT-1997) Yoshito Sadale, National Institute of  
Genetics, Radiolotope Center, Yata 111, Mishima, Shizuoka 411,  
Japan (E-mail:ysadale@lab.nig.ac.jp, Tel:81-0559-81-6870,  
Fax:81-0559-81-6870)

FEATURES  
source Location/Qualifiers

1. 16585  
/organism="Bacillus subtilis"  
/strain="Marburg 168"  
/db\_xref="taxon:1423"  
/map="48 degree"  
1. .19  
/gene="gutB"  
<1. .19  
/gene="gutB"  
/EC\_number="1.1.1.14"  
/note="L-Idiotol 2-dehydrogenase"  
/codon\_start=2  
/transl\_table=11  
/product="sorbitol dehydrogenase"  
/protein\_id="BAA22869.1"  
/db\_xref="GI:4432758"  
/translation="VYPNR"  
82. 1484  
/gene="guta"  
82. .88  
/gene="guta"  
93. 1484  
/gene="guta"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="sugar transport protein"  
/protein\_id="BAA22759.1"  
/db\_xref="GI:2522008"  
/translation="MIGTEKYALKAHOSKSTGLSWERIGYGFQDMSYNIIFQV  
NAVLLFYTDVGGIQPAVATLFLVVRVLDALFDPIMLGLDKTNRWCKARPYLLMW  
AFPEALFTFLCFTTPHFGEGNMVYAYVTYILGMSFMQTI PVNSLTGRMTNSVEER  
TVLITRMLLVYIGILSISCATPLATAIIGEDQAFQMTALIIAASIVLNLFSFF  
TVRERIQPKRKKGIGIKTISLVLEKNKPLMLISSFLAIGFNIKLSTMYIFTYNY  
NHKEFVEMGTVLFFGAALISNLEFIPESEKMGKQVMIITAAISLISYAGLHFTPYSS  
IPLFIWLFASGEFTTPLNLAWGVADCVDAEWKGTIRADGVVISMSFINKLQVA  
LAGSFAIYIGIAGYVANTDQTVASLNAIKNMNALIPGFILLSIILIAFYPLTEKRY  
KHIISELEQRSAR"

gene  
RBS  
CDS

gene  
RBS

CDS

1579. .2541  
/gene="fruc"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="sugar transport protein"  
/protein\_id="BAA22760.1"  
/db\_xref="GI:2522009"  
/translation="MRSHSVVCIGELLIDFECTDVIDVDLMGROFLKSAGAPANVSA  
AIAKIGDAAFSQKVGKDPFGYELKRTLDAVHVDTSMLVNDKAPTTLAFVSLKONGE  
RDFVFNRGADALFTLEDIDQELNEAKILHFGSATALLSDPFCAYLRIMSIKDNQ  
FISFDPNVRREDLMRGRVSEFVSAAKAIAYSDFKVSDELEIISGVKDHEKVAIILH  
EIGANIYAVTLGKSGTLLSNGKDREIIPSIPTSIDSTGAGDAFVGAALYOLANTDQI  
QSVADPFVKLREIIVAFANKVCAIVCTKIGALDALPSLDEIGVSL"

RBS  
gene  
CDS

2722. .2727  
/gene="ydfJ"  
2722. .3420  
/gene="ydfJ"  
2737. .3420  
/gene="ydfJ"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="membrane protein"  
/protein\_id="BAA22761.1"  
/db\_xref="GI:2522010"  
/translation="MSIIGRFKDIMSANINALDKAENPEKNVDOYLRMNSDLAKVK  
AETAAVMAEQRAKREYHECQADMEKESYAMKALQAGNESARKFLERKTSLESKLS  
ELQANQIAATNAAQMRKMHDKLVSDIGELARKNMIKAKVAVAKTQERMNKLGAHSV  
STQSMSAFCGRMEDKVNKALDQANAMAEELNSAPQDDMADLSAKYDTGSSQVDELAA  
LKAKMMLDK"  
join(3472. .3479,3486. .4511)  
/gene="ydfG"  
3472. .3479  
/gene="ydfG"  
3486. .4511  
/gene="ydfG"  
/note="function unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAA22762.1"  
/db\_xref="GI:2522011"  
/translation="MIIISKPCPCGSDMAFDSETGSLSCSCGRODNIESLPKENIAA  
RFSDDEAKEYOCKNCGAVLITFEATATTCSFCGGAITLADRLSGHLAPAKVIPPTIS  
KQEAQOAFRKWKCKGGLTPRGFSADRIKSITGMYIPFWMFDLNSEVOVRANCTRVHQ  
VEEGDYICTETEHEFAFRDINDLYLKIPVDASEKMKDELMDKLEPYSEELKDFQRAY  
LAGTIAEKYNTDEELFPRAKEIISYIDSYIHSTFGTYSVNVVDKHIHTKNVNSFY  
VLPLVMNVSYDYERAHEHIFAMNQDTGKVVGKPPISRKVAAMFSGIAGCTFLAKLVS  
LMGCGF"

gene  
RBS  
CDS

RBS  
gene  
gene  
CDS

gene  
RBS  
CDS

4497. .4501  
/gene="ydfH"  
4497. .4501  
/gene="ydfH"  
4511. .5275  
/gene="ydfH"  
4511. .5275  
/gene="ydfH"  
/note="function unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAA22763.1"  
/db\_xref="GI:2522012"  
/translation="MRGFGKATFVVLAVFIMPLLGIEAVRASELOQHVVYDRAHLIS  
KAEIGKLESIAKLAGAKRDTDFIITTKSTNGEDIADYTGDFYDRYKGSTAILPTIDM  
TNREYFIAGFKKAOYLDNSRLNSIRNTISSDLSENENFEAFETYIQLSYKDMGIRPG  
VNPDNIEFTTWFFQILAAIYAVGTAIVSIMLYHAGGKVTVNGSTYMDQRTSDVIDQYDTY  
IRFTYTRERKPSDKDSGDGVTYKGTSGSRGSF"  
5296. .6277  
/gene="ydfI"  
5296. .5301  
/gene="ydfI"  
5306. .6277  
/gene="ydfI"



```

/note="function unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA22764.1"
/db_xref="GI:2522013"
/translation="MSFRRNQLANVWEEFRDMMFYKWNREIKKSGRLLIRPGD
AVFLNGRVEGVFQDDGDYDIESEIIPFLSTLKGFKFGNSGMRAEVLFTNTKEFTVR
WGTQALINIPAGMPGMPPIRANGTFNFKVDYISLIDKTAGVKDQYFVEDIKRITS
ILDPLMKWITREGKDMFNQANAFDIAGIOEDLDMDMTVTGFOIMSFPNQ
EVQDMITKNASGYGMVDVNRQOISMTDGMASGKMSGSAASDMAGMMGMNMANQMM
NOMNQOAOSSGPOSTGSGSKPNCPCNGTKTGEANFCPNCQKLV"
6291..6323
/note="stem-loop"
/complement(6324..7365)
/gene="ydlj"
/complement(6324..7349)
/gene="ydlj"
/note="function unknown"
/codon_start=1
/transl_table=11
/protein_id="BAA22765.1"
/db_xref="GI:2522014"
/translation="MNEGIITAGLLLTAGMIDFLWTITLESAGPITRCLSAWLK
GCRKISGDHAKVLSMAGPLLCITLVITWISLFWSGWVLVSSDPHSLMETOSKEPASW
SDRIYSGYVMTFLNGDLNAPGMLKLVITETAGCLTITFESVTLISVLSAVNQK
RSFAGSVLSIGMDGETIVHNAMGKDFHDIDFLVAASSELGKLTQHNAPILHYH
STOHQESSIIAVAVLDALTYKYGIPEOYQPNQLIKERSSIKNYLDTVHTAYTHP
AEQAPPEPDISKLQOSGIPALSKOTFOIAVNSIKERRQLLIGITQAGARKWPVEQAI
GNAYSPK"
complement(7358..7365)
/gene="ydlj"
7922..9355
/gene="ydlj"
7922..7927
/gene="ydlj"
7934..9355
/gene="ydlj"
/note="putative"
/codon_start=1
/transl_table=11
/product="metabolite transport protein"
/protein_id="BAA22766.1"
/db_xref="GI:2522015"
/translation="MNKQGNMSPFLRTIILVSTFGGLFGYDTGVNLGALPYMGEPDQ
LNLNATEGLVTSSLFGAALGAVGGRMSDFNGRKNIFLAVIFISTIGCTFAPN
VTVMISREVLGIAVGASVTPAYLAEMSPVESRGVATONELMIVSGQLAFYENA
ILGTMGDNSHWRFMLVIASLPALFLFEGMIRMPESPRLVSKRKEDALRYLKIR
DEKRAAELOEIEFAFKKEDOLEKATFKDLSVPWVRRIYFGLGIAIVQOITGVNSIM
YGTETILRNSGEFOTEAALIGNANGVISLATFVGILGRVGRPMLMTGLIGTTA
LLIGISLVLEGGSPALPYVLSLTVTFIAFOQGAISPVTWMLSEIFPLRLRGIMG
VTVECLMNVNEAVSETPELILAIIGLSTTFIEVGLICSVLEFKRFLPETKGLSLED
LEENFRAYDHSKAKKDSGAEVIG"
9360..9393
/note="stem-loop"
/complement(9403..10457)
/gene="ydl"
/complement(9403..10443)
/gene="ydl"
/note="putative"
/codon_start=1
/transl_table=11

alignment_scores:
  Quality: 87.00      Length: 117
  Ratio: 1.526      Gaps: 5
  Percent Similarity: 48.718      Percent Identity: 26.496

alignment_block:
US-09-728-670-10 x AB007638
Align seg 1/1 to: AB007638 from: 1 to: 16585
5 AsplysGlyLysTyrLysLysGlyAspAspAlaSerTyrpHecLupProth 21

```

```

|||||  ::::|  ||::|
11397 GATAAAGGC...CATGAGAAGGCTGCTGATGTTTCAGTCAATACCGAAG 11443
21 rglpProTyrLeuMetValAsnValThrGlyValAspSerLysGlyAsnG 38
   |||  ||::|  ::::|
11444 CGGTGATGAGCTGCTGCTTTCATATACAGATACGAT... 11480
38 luleuLeuSerProHisTyrValGluPheProIleLysProGlyThr 54
   |||||  ::::|  ::::|
11481 .CTGCTGACAAAGTATTAT.....GAAATGACAAAGTG 11513
55 leuThrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAspAlaTh 71
   ::  ::::|  ::::|  ||  ::::|  ||
11514 ATCCATGAGAGAAGAGCTGACCAGCATATCCGCATTTGCTTGAT.... 11558
71 rAlaTyrLysGluPheArgValValGluLeuAspProSerAlaLysIleG 88
11559 .....CAAAAACAGC 11568
88 luValThrTyrTyrAspLysAsnLysLysGluGluThrLysSerPhe 104
   ::||  |||||  ||::|  ::::|  ::::|  ::::|
11569 AGGTGCTCTATTATACGGCAATATGAACAGATGAATGAGGCTTTT 11618
105 proIleThrGluLysGlyPheValValProAspLeuSerGluHisIlely 121
   ::  |||  ::::|  ::::|  ::::|
11619 AAATTG.....CATCTGAAGTCTCATATAAAA 11644
121 s 121
11645 A 11645

seq_name: gb_ba:BSUB0004

seq_documentation_block:
LOCUS      BSUB0004      213190 bp      DNA      linear      BCT 26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 4 of 21): from 600701 to
81890.
ACCESSION  Z99107 AL009126
VERSION    Z99107.1 GI:2632866
KEYWORDS
SOURCE
ORGANISM   Bacillus subtilis.
            Bacillus subtilis
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE
AUTHORS    Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertoletti,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriess,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Brusch,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Conneron,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emerson,P.T., Entlan,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guisepi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Konigstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Maue,D., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Portwollik,S., Prescott,A.M., Prescan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadale,Y., Sato,T.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpsira,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Welfzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,

```

TITLE Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and Danchin, A.  
The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*  
JOURNAL Nature 390 (6657), 249-256 (1997)  
MEDLINE 98044033  
REFERENCE 2 (bases 1 to 213190)  
AUTHORS Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.  
TITLE Direct Submission  
JOURNAL Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

FEATURES  
source location/Qualifiers  
1. 213190  
/organism="Bacillus subtilis"  
/strain="168"  
/db\_xref="taxon:1423"  
33. .740  
/gene="ydfg"  
33. .740  
/gene="ydfs"  
/function="unknown"  
/note="similar to hypothetical proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12373.1"  
/db\_xref="GI:2632867"  
/db\_xref="SPTREMBL:P96697"  
/translation="MIELEVIVRTVASFGLLIARILGKOTISOMTIDFIAITLG AIAAGLAYNTSIRKPHNMAISFIVLTIFLISFISIKNRKLKFFAGDPVLIQNGKI LESNMKRMRYTLDYLDNQREKEIFNIEVLFALETNGOLTVRKPOFRVTKODLM IAVNQEQRPLIELMDGEIIEENLKQNRLTESWLEELRKRDIKVKEIVAVLLGNGD IYVDQYKDHISVPMDKE"  
gene complement(755. .1186)  
/gene="ydfg"  
CDS complement(755. .1186)  
/gene="ydfg"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12374.1"  
/db\_xref="GI:2632868"  
/db\_xref="SPTREMBL:P96698"  
/translation="MDFEIRKMLEITNEYKQSDFTNVLYKKAPENHFFDSEASTFVY DRYQDEEYNFITIVEMPYVEEELTIRLSKTQLKITTPVFPAEMEVLKERYGGEI ERIQLPEAAETHLQIQLNLHISYPROVEIVAFNKGK"  
gene complement(1199. .1441)  
/gene="ydgA"  
CDS complement(1199. .1441)  
/gene="ydgA"  
/function="unknown"  
/note="similar to hypothetical proteins from B. subtilis"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12375.1"  
/db\_xref="GI:2632869"  
/db\_xref="SPTREMBL:P96699"  
/translation="MPYQINIANIKINGVTQNGNIDVGPVHNSHTANSKYFGANFSL GDLSPSSLLNTGNIDSDVDQDQIGNPAPISNQI"  
gene complement(1455. .1727)  
/gene="ydgB"  
CDS complement(1455. .1727)  
/gene="ydgB"  
/function="unknown"  
/note="similar to hypothetical proteins from B. subtilis"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12376.1"  
/db\_xref="GI:2632870"  
/db\_xref="SPTREMBL:P96700"  
/translation="MPSFVINLYYLKINSISGNSITIGEAAVNSPTNNQKSGGTNSS

gene EGDTPTESVMENFLNDPDVNDQTSIGNSDTSNVNAPPIAPPILD"  
2026. .2613  
/gene="ydgC"  
CDS 2026. .2613  
/gene="ydgC"  
/function="unknown"  
/note="similar to hypothetical proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12377.1"  
/db\_xref="GI:2632871"  
/db\_xref="SPTREMBL:P96701"  
/translation="MKSQGVESRRLKAANEFSGVGFHDQAVSEIYKAGFTQPSF YLYFQSKFAIFAELITDFHSRVKRLTESLLENGLTEDYSKRVLLAVETVFOPLDED KDLTKIGFELNPEAKOMKDLAMVLEKNELEAEQRLGFHSELDMETVAECLVGMIEHL TESFLTGIKDPASLAEVNVLLIYGLPKGNDVR"  
gene join(2610. .2954,3074. .3091)  
/gene="ydgD"  
CDS 2610. .2954  
/gene="ydgD"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12378.1"  
/db\_xref="GI:2632872"  
/db\_xref="SPTREMBL:P96702"  
/translation="MISIMMKVSLAVFMLAGGIKVSVPQVEHWRHYQPLMELTV TGILEIAGLAMTAGIWNRYAALGAGVLFVLMAGAIHMHMFRARQSVIMAIQMICL IVSIMIMGSYT"  
terminator 3074. .3091  
/gene="ydgD"  
gene 3117. .3590  
/gene="ydgE"  
CDS 3117. .3590  
/gene="ydgE"  
/function="ydgE"  
/function="unknown"  
/note="similar to hypothetical proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12379.1"  
/db\_xref="GI:2632873"  
/db\_xref="SPTREMBL:P96703"  
/translation="MAFTLEDMTTEEFEEAFRGMVSQNFPAKONITSGTWTEKEAFEXSE QAYENMIPNGRDSNNHYFWNITFNDGERMGWLMVADPLHPQKEAFIYSGLYEAFRG KGLAQLALQTLDERARELGAERLALHVFANHETAVVLYQKGYAMTNIRMRKQLC"  
gene complement(3705. .3722)  
/gene="expZ"  
CDS complement(3705. .5393)  
/gene="expZ"  
complement(3750. .5393)  
/gene="expZ"  
/codon\_start=1  
/transl\_table=11  
/product="ATP-binding transport protein"  
/protein\_id="CAB12380.1"  
/db\_xref="GI:2632874"  
/db\_xref="SWISS-PROT:P39115"  
/translation="MKEIVTLTNVSYEVKQDTVEKHNASVQGDIIIGIKNGAGKS TLHLIHNDLAPAQOQILRKDIKALAVEQETAAYSFADQTPAEKKLLEKWHVPLDFH QLSGGEKLRARLAKGLEDADLLDEPTNHLDEKSLQFLQOLKHYNGTVILVSHDR YFLDEAATRIWSLEDQTLIEFKGNSGYMKFREKKRLTQOREYEKQOKMVERIEAOMN GLAWSSEKAHAQSTKKEGFEKHYRVAKRTDAOIKSKQRLKELEKAKAEVPPEYT VRESIDTTHKTGKRFLLEVQNVTKAFGERTLFKNANFTIQGEKVAIIGPNSGKTTLL NIILGOETAGESVWVSPSANIGYLTQVEVDLPLEQTPPELFEENETFKARGHYONLMRH LGFTAAQWTEPIKHMMSMERKIKLMAYILLEKDVLLDEPTNHLDLPSRQLEETLS QYSGTLAVSHDRYFLKETNSKLVISNNGIEKQLVNDVPSERNEREELRLKLETGEQE VLKGLSEFMTPNDKGYKELDQAFNELTKRIKELDHQDKKD"  
terminator complement(5428. .5450)  
/gene="ydgF"  
gene complement(5428. .7089)  
/gene="ydgF"  
terminator complement(5678. .5703)  
/gene="ydgF"

```

CDS
complement(5713..7089)
/feature="ydgf"
/function="unknown"
/notes="similar to amino acid ABC transporter (permease)"
/codon_start=1
/translation="1"
/translation="MTDDMTKDNINQQLQGLKNRHIQLAIGGAICTGLFLGSGKS
IHFGPSILFAYMITGIIICFLIMRSDELLSNLNSHFDVFDYDGMAAFITGWT
YWFCTISIAMADLVAGLTYQYWLPGYPOWPGILALIILLIMNLATYKLFGELEFW
ALIKVIALIALIVIGLVMIFKGFSTSGVSSFTNLMSHGLFPNGMGFILSFOMVF
AFVGIELVGLTAGETENPEKVIKPAINIPVRVLLFYIGALLVIMSIYPMDIINPSES
PFVQVEVAVGIVGAASIINFVLTSAASACNSAVFSTSRMVYSLAKDHNAPESMAKLT

```

```

FEATURES             Maxygen Aps (DK)
                      Location/Qualifiers
                      1. .66
                      /organism="synthetic construct"
                      /db_xref="taxon:32630"
                      /note="Primer"
BASE COUNT          16 a      11 c      9 g      30 t
ORIGIN
alignment_scores:
    Quality:      86.00      Length:      18
    Ratio:        5.059      Gaps:      0
Percent Similarity: 94.444      Percent Identity: 94.444

alignment_block:
US-09-728-670-10 x AX073718/rev ..

Align seg 1/1 to reverse of: AX073718 from: 1 to: 66

119 HisilelysAsnProGlyPheAsnLeuileThrLysValValilegIuLy 135
|||||
65 CATATTAAAAACCCTGGAATTCAACTTAAGACAAAGTTGTATAGAAAA 16

135 slys 136
||||
15 GAAA 12

seq_name: gb_pat:AX073715

seq_documentation_block:
LOCUS      AX073715                      66 bp      DNA      linear
DEFINITION Sequence 8 from Patent WO0104287.
ACCESSION  AX073715
VERSION     AX073715.1  GI:12710133
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS     Halkier,T., Pedersen,A.H. and Okkels,J.S.
TITLE       A method for preparing modified polypeptides
JOURNAL     Patent: WO 0104287-A 8 18-JAN-2001;
            Maxygen Aps (DK)
FEATURES
source
1. .66
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Primer"
BASE COUNT  18 a      9 c      9 g      30 t
ORIGIN
alignment_scores:
    Quality:      85.00      Length:      18
    Ratio:        5.000      Gaps:      0
Percent Similarity: 94.444      Percent Identity: 94.444

alignment_block:
US-09-728-670-10 x AX073715/rev ..

Align seg 1/1 to reverse of: AX073715 from: 1 to: 66

119 HisilelysAsnProGlyPheAsnLeuileThrLysValValilegIuLy 135
|||||
65 CATATTAAAAACCCTGGAATTCAACTTAATACAAAGGTTGTATAGAAAA 16

135 slys 136
||||
15 GAAA 12

seq_name: gb_pat:A059669

```

## seq\_documentation\_block:

LOCUS A05969 11225 bp DNA linear PAT 11-JUN-1993

DEFINITION Complete sequence of the pmh23 insert (the '11.2kb BamHI fragment').

ACCESSION A05969

VERSION A05969.1 GI:412654

KEYWORDS

SOURCE synthetic construct.

ORGANISM synthetic construct

REFERENCE 1 (bases 1 to 11225)

AUTHORS

TITLE FOWLPOX VIRUS PROMOTERS

JOURNAL Patent: WO 8903879-A 1 05-MAY-1989;

FEATURES Location/Qualifiers

1..11225

/organism="synthetic construct"

/db\_xref="taxon:32630"

complement(4821..5138)

/gene="ORF5"

complement(4821..5138)

/gene="ORF5"

/codon\_start=1

/transl\_table=1

/protein\_id="CAA00470.1"

/db\_xref="GI:412655"

/translation="MIRRNKALGSVMSDFIKTINEEYDSNIKEIKSEIDIKNSIL

KELDEKRYROEIKELCMIVDLKQYKIIDNIYSRYITEIRQLLALKEENKCKEELT

KLK"

8025..8375

/gene="ORF8"

8025..8375

/gene="ORF8"

/codon\_start=1

/transl\_table=1

/protein\_id="CAA00471.1"

/db\_xref="GI:412656"

/translation="MEEGKPRSSAVLMLIPCGSIIVSVFVILSTRPPVPPDIK

ILYCKEGWYGNKNCYFFSEKKNKSLAVERCKMDGHLTSSISKEEFKILRYKPG

NHWIGIEKVDENGT"

complement(8844..9686)

/gene="ORF10"

complement(8844..9686)

/gene="ORF10"

/codon\_start=1

/transl\_table=1

/protein\_id="CAA00472.1"

/db\_xref="GI:412657"

/translation="MKEKEVRNTIKKMNITDIKICIGINEYFMSMKLLDVEVIMRNG

FVNITRLCNLEKDFNDWKQLESRRLLNTLKDNNKLHDPINIRTRIKINGEYVSQ

LLLDYVIFWISPYVATRVSLMRYRCVALNIETEKDIDHSQELQNIQISKIDEYDR

SIKDISNREKEIETSYSKSLTYLLTKAERVLKDYSMEDIDNNEDIRTDGMTAIE

AEIEENRRYLSIISGIRKQHAEDRINISKIMLSGDSFNEIIVKIRDYIETTAKPAVA

NNE"

BASE COUNT 3669 a 1720 c 1642 g 4193 t 1 others

ORIGIN

## alignment\_scores:

Quality: 84.50

Ratio: 1.408

Percent Similarity: 48.780

Length: 123

Gaps: 5

Percent Identity: 25.203

## alignment\_block:

US-09-728-670-10 x A05969/rev ..

Align seg 1/1 to reverse of: A05969 from: 1 to: 11225

```

43 HistyValGluPheProIleLysProGlyThrThrLeuThrLysGluLys 59
||||| ||||| ::::: ||::: ||::: ||
1113 CATTAATTGTTGAAACCTATACATATAGCTGCTAATAGAAAGAGTAA 11064
59 site...GluTyrTyrValGluTrpAlaLeuAspAlaThrAla..... 72

```

```

|||| :::: ::::: ||| :::::
11063 AATAGTAAATTCCTTATAGAAATACGCCGCTGACATTAATCAAGAAGACG 11014
73 .....TyrlYsglu 75
11013 GCGGAATGCTAAATACCTATACATTAACGCTATGAAGATATACGATCCG 10964
76 PheArgVal.....ValGluLeuAspProSerAlaIleLeu 88
10963 TTTAGATTAATAATAATAAAGTATTAATAGACACGCCGCC..... 10922
88 uValThrTyrTyrAspLysAsnLysLysLysGluGluThrLysSerPhe 105
10921 .....GATATTAACAACAAGCGCTTTTAACATAATACATCCC 10885
105 rolIethrGluLysGlyPheValValProAspLeuSerGluHisIleLys 121
10884 CCTTATACGAACCTAGGTTTATTAACGACGACCTATATAGATTAACATCATA 10835
122 AsnProGlyPheAsnLeu.....Ileth 129
10834 TCTAGAGAGCTAATATATAATAATAAAGAAAGATGGTAGAATATATTAC 10785
129 rLysValValIleGluLys 135
10784 ACGAATAATATTAGAACGA 10766

```



---

100



17 rphegluProthrGlyProTyrLeuMetValAsnValThrGlyValAsps 34  
|||||  
51 TTTGAACCAACAGAGCCGATTTGATGTAATGTGACTGGAGTTGATA 100  
34 erlysgLyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
|||||  
101 GTAAGGAATGAATTGCTATCCCTCATTTATGTCGAGTTTCCCTATTAAA 150  
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrpaI 67  
|||||  
151 CTTGGGACTACACTTACAAAAGAAAATGTAATCTATGTGCAATGGGC 200  
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84  
|||||  
201 ATTAGATGCGACAGCATATAAGAGTTTAGAGTAGTTGAATTAGATCCAA 250  
84 eraLalysIleGluValThrTyrTyrAspLysAsnLysLysLysGluGlu 100  
|||||  
251 GCGCAAGATCGAAGTCACCTTATTATGATTAAGATAAGAAAAAGAGAA 300  
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117  
|||||  
301 ACGAAGTCTTCCCTATAACAGAAAAAGTTTGTGTCCAGATTATC 350  
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134  
|||||  
351 AGAGCATATTAAAAACCCTGATTCACACTTAATTACAAGAGTTGTATTAG 400  
134 LuLysLys 136  
|||||  
401 AAAAGAAA 408

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:ABA05083

seq\_documentation\_block:

ID ABA05083 standard; DNA; 411 BP.

XX ABA05083;

DT 22-FEB-2002 (first entry)

DE Human dipeptide aminopeptidase 28 coding sequence.

XX Human; dipeptide aminopeptidase 28; cancer; nosohaemia; cytostatic;  
KW anti-HIV; immunosuppressive; antiinflammatory; HIV infection;  
KW immunological disease; inflammation; gene therapy; ds.

XX Homo sapiens.

Key Location/Qualifiers

FT CDS 1..411

FT /tag- a

FT /product- "dipeptide aminopeptidase 28"

FT /partial

FT /note- "the sequence contains no start codon"

PN CN1307129-A.

XX 08-AUG-2001.

PF 03-FEB-2000; 2000CN-0112673.

XX 03-FEB-2000; 2000CN-0112673.

PA (JINP-) JINPENG BIO TECH CO LTD SHANGHAI.

PI Wang J, Wang S, Tong B;

DR WPI; 2002-034896/05.

DR P-PSDB; AAM47321.

PT New polypeptide for treating malignant tumors and HIV infection,  
comprises the human dipeptide aminopeptidase -

XX Claim 8; page 1-2(Claims); 11pp; Chinese.

XX The present invention provides the protein and coding sequences of human  
CC dipeptide aminopeptidase 28. The sequences can be used in the treatment  
CC of cancer, nosohaemia, HIV infection, immunological diseases and  
CC inflammation. The present sequence is the coding sequence of the  
CC invention.

XX Sequence 411 BP; 162 A; 59 C; 80 G; 110 T; 0 other;

alignment\_scores:

Quality: 715.00 Length: 136

Ratio: 5.257 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-728-670-10 x ABA05083 ..

Align seg 1/1 to: ABA05083 from: 1 to: 411

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTy 17

1 TCAGTTTCATTTCGACAAAGGAATAATAAAAAAGCCGATGACGCGAGTTA 50

17 rphegluProthrGlyProTyrLeuMetValAsnValThrGlyValAsps 34

51 TTTTGAACCAACAGAGCCCGTATTGATGTAATGTGACTGGAGTTGATA 100

34 erlysgLyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50

101 GTAAGGAATGAATTGCTATCCCTCATTTATGTCGAGTTTCCCTATTAAA 150

51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrpaI 67

151 CTTGGGACTACACTTACAAAAGAAAATGGAATACATGTCGAATGGGC 200

67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84

201 ATTAGATGCGACAGCATATAAGAGTTTAGAGTAGTTGAATTAGATCCAA 250

84 eraLalysIleGluValThrTyrTyrAspLysAsnLysLysLysGluGlu 100

251 GCGCAAGATCGAAGTCACCTTATTATGATAAGATAAGAAAAAGAGAA 300

101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117

301 ACGAAGTCTTCCCTATAACAGAAAAGGTTTGTGTGTCCAGATTATC 350

117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134

351 AGAGCATATTAAAAACCCTGATTCACACTTAATTACAAAGTTGTATTAG 400

134 LuLysLys 136

401 AAAAGAAA 408

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44270

seq\_documentation\_block:

ID AAQ44270 standard; DNA; 414 BP.

XX AAQ44270;

DT 03-DEC-1993 (first entry)

DE Staphylokinase SAK-STAR.

KW Signal sequence, plasminogen activator; thrombosis; staphylokinase;

KW SAK; ss.

OS Staphylococcus aureus strain 23.



```

XX  WO9313209-A.
PN
XX  08-JUL-1993.
PD
XX
XX  28-DEC-1992; 92WO-EP02989.
PE
XX
XX  30-DEC-1991; 91DE-4143279.
PR  22-JUN-1992; 92DE-4220516.
PR  01-DEC-1992; 92DE-4240801.
XX
XX  (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.
XX
XX  Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
PI
XX  WPI; 1993-227325/28.
DR  P-PSDB; AAR39150.
DR
XX
XX  Staphylo-kinase (SAK) sequences lacking sequences for signal
PT  peptide(s) - for prodn. of proteins used as plasminogen
PT  activators in thrombosis treatment, and monoclonal antibodies
PT  against SAK
XX
XX  Claim 3; Fig 3; 99pp; German.
PS
XX  DNA encoding SAK lacking the signal peptide, is expressed
CC  intracellularly. This avoids the problem of fast degradation of the
CC  polypeptides or destruction of the host when expressed into the
CC  medium or into the periplasm respectively. High expression is
CC  possible and the chemically induced overprod. is easy to handle.
CC  Also, the prods. are homogeneous.
CC  SAK-polypeptide derivs. are plasminogen activators for the
CC  treatment of thrombosis.
XX
SQ  Sequence 414 BP; 163 A; 59 C; 81 G; 111 T; 0 other;

```

```

alignment_scores:
      Quality: 715.00      Length: 136
      Ratio: 5.257        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment\_block:

us-09-728-670-10 x AAQ44270 ..

Align seg 1/1 to: AAQ44270 from: 1 to: 414

```

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
   |||||||
4 TCAAGTTTCATTCGACAAAGGAAATATATAAAAAAGCGATGACGCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
   |||||||
54 TTTTGAACCAACAGGCCCGTATTGTGATGTAATGTGACTGAGTTGATA 103
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
   |||||||
104 GTAAAGGAAATGAATGCTATCCCTCATTTATGTCGAGTTCTTATAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTPrAl 67
   |||||||
154 CCTGGACTACACTTACAAAGAAATAATGAATACTATGTCGAATGGGC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
   |||||||
204 ATTAGATCGCACAGCATATAAAGAGTTTAGAGTAGTTGAATTAGATCCA 253
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
   |||||||
254 GCGCAAGATCGAAGTCACTTATATGATTAAGAATTAAGAAAAAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
   |||||||

```

```

304 ACGAAGTCTTCCCTATACAGAAAAAGCTTTGTGTGCCAGATTATTC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
   |||||||
354 AGACATATTAAACCCCTGATTCAACTTAATTACAAAGGTTGTATAG 403
134 LuLysLys 136
   |||||||
404 AAAAGAAA 411

```

seq\_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44269

seq\_documentation\_block:

ID AAQ44269 standard; DNA; 414 BP.

XX AC AAQ44269;

DT 03-DEC-1993 (first entry)

XX DE Staphylokinase SAK-Cphic.

XX KW Signal sequence, plasminogen activator; thrombosis; staphylokinase;

KW SAK; ss.

XX OS Staphylococcus aureus phage phiC.

XX PN WO9313209-A.

PD 08-JUL-1993.

XX PF 28-DEC-1992; 92WO-EP02989.

XX PR 30-DEC-1991; 91DE-4143279.

XX PR 22-JUN-1992; 92DE-4220516.

XX PR 01-DEC-1992; 92DE-4240801.

XX PA (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.

XX PI Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;

XX DR WPI; 1993-227325/28.

XX DR P-PSDB; AAR39149.

XX PT Staphylo-kinase (SAK) sequences lacking sequences for signal

XX PT peptide(s) - for prodn. of proteins used as plasminogen

XX PT activators in thrombosis treatment, and monoclonal antibodies

XX PT against SAK

XX PS Claim 3; Fig 2; 99pp; German.

XX CC DNA encoding SAK lacking the signal peptide, is expressed

XX CC intracellularly. This avoids the problem of fast degradation of the

XX CC polypeptides or destruction of the host when expressed into the

XX CC medium or into the periplasm respectively. High expression is

XX CC possible and the chemically induced overprod. is easy to handle.

XX CC Also, the prods. are homogeneous.

XX CC SAK-polypeptide derivs. are plasminogen activators for the

XX CC treatment of thrombosis.

SQ Sequence 414 BP; 161 A; 59 C; 83 G; 111 T; 0 other;

alignment\_scores:

```

      Quality: 711.00      Length: 136
      Ratio: 5.228        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.265

```

alignment\_block:

us-09-728-670-10 x AAQ44269 ..

Align seg 1/1 to: AAQ44269 from: 1 to: 414

```

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17

```

```
|||||
4 TCAGTTCATTCGACAAAGAAATATATAAAGCGCATGACGAGTTA 53
17 rphgluprothrGlyProtyrLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTTTGAACCAACACAGCGCCGATTTGATGTAATGCTGAGTGTGATG 103
34 erlysglyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::
104 GTAAAGGAATGAATGCTATCCCTCATATATGTCGAGTTCCCTATAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
|||||
154 CCTGGGACTACACTTACAAAGAAATAATGAATACTATGTCGAATGGCC 203
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
204 ATTAGATGCGACAGCATATTAAGAGTTTAGAGTAGTTGAATTAGATCCA 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
254 GCGCAAGATCGAAGTCACTTATGATAGAATAAGAAAAAGAAAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
|||||
304 ACGAAGTCTTCCCTATTAACAGAAAAAGTTTGTGTCCTCCAGATTATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
|||||
354 AGAGCATATTAATAAACCTGATTCACCTTAATTACAAAGCTTGTATAG 403
134 LuLysLys 136
|||||
404 AAAAGAAA 411

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:AAQ11813
seq_documentation_block:
ID AAQ11813 standard; DNA; 804 BP.
AC AAQ11813:
XX
XX 05-AUG-1991 (first entry)
DT
XX
DE Staphylococcus aureus promoter, Shine-Dalgarno and signal peptide
DE sequences.
XX
XX somatomedin; protein production; ds.
KW
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FH sig_peptide 313..393
FT /*tag= a
FT promoter 193..222
FT /*tag= b
FT RBS 301..305
FT /*tag= c
FT /standard_name= Shine-Dalgarno sequence
XX
XX JP03098595-A.
XX
XX 24-APR-1991.
PD
XX
XX 11-SEP-1989; 89JP-0234874.
PF
XX
XX 11-SEP-1989; 89JP-0234874.
PR
XX
XX (TAIS ) TAISHO PHARMACEUT KK.
PA
XX
XX WPI; 1991-167039/23.
DR
DR P-PSDB; AAR12137.
XX
```

```
PT Prepn. of peptide(s) - by construction of expression
PT vector, transformation of E.coli etc., culturing to secrete
PT peptide(s) and collecting peptide(s)
XX
XX Example; Fig 3; 15pp; Japanese.
PS
XX
XX The sequence up to the EcoRI site at position 544 is ligated to a
CC synthetic Somatomedin C coding sequence to give a SAK-SMC construct.
CC Bacillus subtilis or E.coli can be transformed with such a
CC recombinant construct for expression and extracellular secretion of
CC somatomedin C. The SMC is secreted as a fusion protein with the
CC Staphylococcus gene product. See also AAQ11814.
XX
SQ Sequence 804 BP; 294 A; 89 C; 148 G; 273 T; 0 other;

alignment_scores:
Quality: 711.00 Length: 136
Ratio: 5.228 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.265

alignment_block:
US-09-728-670-10 x AAQ11813 ..

Align seg 1/1 to: AAQ11813 from: 1 to: 804

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
394 TCAGTTCATTCGACAAAGAAATATATAAAGCGCATGACGAGTTA 443
17 rphgluprothrGlyProtyrLeuMetValAsnValThrGlyValAsp 34
|||||
444 TTTTGAACCAACAGCGCCGATTTGATGTAATGTGACTGAGTGTGATG 493
34 erlysglyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::
494 GTAAAGGAATGAATGCTATCCCTCATATATGTCGAGTTCCCTATTAA 543
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
|||||
544 CCTGGGACTACACTTACAAAGAAAAATGAATACTATGTCGAATGGGC 593
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
594 ATTAGATGCGACAGCATATTAAGAGTTTAGAGTAGTTGAATTAGATCCA 643
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
644 GCGCAAGATCGAAGTCACTTATGATAGAATAAGAAAAAGAAAGAA 693
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
|||||
694 ACGAAGTCTTCCCTATTAACAGAAAAAGTTTGTGTCCTCCAGATTATC 743
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
|||||
744 AGAGCATATTAATAAACCTGATTCACCTTAATTACAAAGCTTGTATAG 793
134 LuLysLys 136
|||||
794 AAAAGAAA 801

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAc85112
seq_documentation_block:
ID AAC85112 standard; DNA; 1377 BP.
XX
XX AAC85112:
AC
XX
XX 08-MAY-2001 (first entry)
DT
XX
DE DNA sequence of S. aureus staphylokinase.
XX
```

KW Immunogenicity; staphylokinase; variant; stability; ds.  
XX Staphylococcus aureus.  
OS  
FH Key Location/Qualifiers  
FT promoter 193..222  
FT RBS /\*tag= a  
FT /\*tag= b  
FT CDS 301..305  
FT /\*tag= b  
FT 313..804  
FT sig\_peptide /\*tag= c  
FT 313..393  
FT /\*tag= d  
FT mat\_peptide 394..801  
FT /\*tag= e  
FT /note= "sequence AAB61908"  
XX  
PN WO200104287-A1.  
XX  
PD 18-JAN-2001.  
XX  
PF 06-JUL-2000; 2000WO-DK00371.  
XX  
PR 07-JUL-1999; 99DK-0000988.  
PR 27-AUG-1999; 99DK-0001196.  
PR 02-MAR-2000; 2000DK-0000339.  
PR 18-MAY-2000; 2000DK-0000804.  
XX  
PA (MAXY-) MAXYGEN APS.  
XX  
PI Halkier T, Pedersen AH, Okkels JS;  
XX  
DR WPI: 2001-138342/14.  
DR P-PSDB; AAB61908.  
XX  
XX  
PT Producing polypeptides with altered immunogenicity or improved  
PT stability, comprises expressing a diversified nucleotide sequence  
PT population and selecting polypeptides with altered immunogenicity or  
PT improved stability -  
XX  
PS Example 1; Page 74-75; 83pp; English.  
XX  
XX The invention relates to a method of altering immunogenicity and/or  
CC increasing stability of a polypeptide of interest. The method comprises  
CC (a) expressing a diversified population of nucleotide sequences encoding  
CC a polypeptide of interest; (b) screening the polypeptides expressed for  
CC function, immunogenicity and/or stability; and (c) selecting functional  
CC polypeptides with altered immunogenicity and/or increased stability. The  
CC method is used to improve the properties of polypeptides, in particular  
CC to alter the immunogenicity and/or increase the functional in vivo half-  
CC life of the polypeptide. The method uses a high throughput system that  
CC makes it possible to search several orders of magnitude more polypeptides  
CC than is possible by previously known approaches. This enhances the chance  
CC of finding the optimal variant from the many thousands of variants that  
CC may be produced. The present sequence represents a DNA encoding a  
CC S. aureus staphylokinase, used to exemplify the method of the invention.  
XX  
SQ Sequence 1377 BP; 452 A; 184 C; 255 G; 486 T; 0 other;

alignment\_scores:  
Quality: 711.00 Length: 136  
Ratio: 5.228 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.265

alignment\_block:  
US-09-728-670-10 x AAC85112 ..  
Align seg 1/1 to: AAC85112 from: 1 to: 1377

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17  
|||||  
394 TCAAGTTCATTCGACAAAGGAAATATATAAAAGGCGATGACGGAGTTA 443

17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
444 TTTTGAACCAACACAGCCCGTATTTGATGTAATGTGACTGAGTTGATG 493  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
:|||||  
494 GTAAAGGAATGAATTGCTATCCCTCATTTATGTGAGTTTCTTATTAA 543  
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67  
|||||  
544 CCTGGACTACACTTACAAAAGAAAATTTGAATACATGTCGAATGGGC 593  
67 aleuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspProS 84  
|||||  
594 ATTAGATGCACACACATATAAAGAGTTTAGAGTAGTTGAATTAGATCCAA 643  
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100  
|||||  
644 GCGCAAGATCGAAGTCACTTATTTAGATAAGATAAGAAAAGAGAA 693  
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117  
|||||  
694 ACGAAGTCTTTCCCTATTAACAGAAAAGTTTGTGTGTCAGATTATTC 743  
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134  
|||||  
744 AGACATATTAATAAACCCCTGATTCACTTAATTACAAAGTTGTTATAG 793  
134 LuLysLys 136  
|||||  
794 AAAAGAA 801

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAx27541

seq\_documentation\_block:  
ID AAX27541 standard; DNA; 408 BP.  
XX  
XX AAX27541;  
XX  
DT 27-MAY-1999 (first entry)  
XX  
XX S. aureus staphylokinase (Sak) encoding DNA.  
DE  
XX Staphylokinase; Sak; recombinant; myocardial infarction; cerebral;  
KW thrombembolia disease; arterial thrombosis; pulmonary thrombosis;  
KW hydrolytic; fibrin; ss.  
XX  
OS Staphylococcus aureus.  
XX  
PN WO9904017-A1.  
XX  
PD 28-JAN-1999.  
XX  
PF 17-JUL-1998; 98WO-CN00129.  
XX  
PR 19-JUL-1997; 97CN-0105988.  
XX  
PA (BAIH/) BAI H.  
XX  
PI Ble L, Qu G, Wu Y, Xu G, Xu W, Zhang G, Zhang Q;  
XX  
DR WPI: 1999-132261/11.  
DR P-PSDB; AAY01232.  
XX  
PT Highly safe, novel recombinant staphylokinase (Sak) produced from  
PT high-expression engineered strain - as plasminogen activator, with  
PT very high hydrolytic activity to human fibrin, useful in treating  
PT thrombembolia diseases e.g. myocardial infarction  
XX  
PS Claim 2; Page 24; 51pp; Chinese.  
XX  
CC This DNA encodes a recombinant staphylokinase (Sak). The Sak-producing

CC Staphylococcus aureus SL1.063 is deposited as CGMC No.0353. The  
 CC invention provides a method for constructing a Sak-producing engineered  
 CC strain which comprises (a) screening Sak gene by PCR (polymerase chain reaction)  
 CC aureus; obtaining a required Sak gene by PCR (polymerase chain reaction)  
 CC amplification with chromosomal DNA of Sak-producing S. aureus as  
 CC template, with primers (AAx27542-43) (b) introduction of the obtained  
 CC DNA fragment into a plasmid selected from pUC19 and pBV220; and (c)  
 CC transferring the recombinant plasmid into a host cell such as E. coli of  
 CC DH5 alpha, TGI or TG2 strains. The staphylokinase can be applied in  
 CC treatment of myocardial infarction, thrombembolia diseases and arterial  
 CC thrombosis including cerebral and pulmonary thrombi. Hydrolytic activity  
 CC of the staphylokinase to human fibrin is very high.  
 XX  
 SQ Sequence 408 BP; 156 A; 59 C; 83 G; 110 T; 0 other;

alignment\_scores:  
 Quality: 703.00 Length: 136  
 Ratio: 5.169 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 98.529

alignment\_block:  
 US-09-728-670-10 x AAx27541 ..

Align seg 1/1 to: AAx27541 from: 1 to: 408

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17
| | | | | | | | | | | | | | | | | | | | | | | |
1 TCAAGTTCATTGACAAAGAAATATATAAAAGCGATGACGCGATT 50
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
| | | | | | | | | | | | | | | | | | | | | | | |
51 TTTGAACCAACAGGCCCGTATTGATGTAATGTGACTGAGTTGATG 100
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
: | | | | | | | | | | | | | | | | | | | | | |
101 GTAAGGAAATGAATGCTATCCCTCGTTATGTCGAGTTCCATTAA 150
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyr 67
| | | | | | | | | | | | | | | | | | | | | | | |
151 CCTGGACTACACTTACAAAGAAATTAATGATATGTCGATGGGC 200
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
| | | | | | | | | | | | | | | | | | | | | | | |
201 ATTAGATGCGACAGCATATATAAGAGTTAGAGTAGTGAATTAGAT 250
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
| | | | | | | | | | | | | | | | | | | | | | | |
251 GCGCAAGATCGAAGTCACTTATTATGATTAAGATAAGAAAAAGAA 300
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
| | | | | | | | | | | | | | | | | | | | | | | |
301 ACGAAGTCTTCCCTATTAACAGAAAAAGTTTGTGTCCAGATTATC 350
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
| | | | | | | | | | | | | | | | | | | | | | | |
351 AGAGCATATTAATAAACCTGATTCACCTTAATTACAAAGTTGTTAT 400
134 LuLysLys 136
| | | | | |
401 AAAAGAAA 408
```

seq\_name: /SID51/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44268

seq\_documentation\_block:  
 ID AAQ44268 standard; DNA; 414 BP.

XX  
 AC AAQ44268;  
 XX  
 DT 03-DEC-1993 (first entry)  
 XX  
 DE Staphylokinase SAK42D.  
 XX

KW Signal sequence, plasminogen activator; thrombosis; staphylokinase;  
 KW SAK; ss.

XX Staphylococcus aureus phage 42D.

OS WO9313209-A.

XX 08-JUL-1993.

XX 28-DEC-1992; 92WO-EP02989.

XX 30-DEC-1991; 91DE-4143279.

XX 22-JUN-1992; 92DE-4220516.

XX 01-DEC-1992; 92DE-4240801.

XX (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.

XX Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;

XX WPI; 1993-227325/28.

XX P-PSDB; AAR39148.

XX Claim 2; Fig 1; 99pp; German.

XX DNA encoding SAK lacking the signal peptide, is expressed

XX intracellularly. This avoids the problem of fast degradation of the

XX polypeptides or destruction of the host when expressed into the

XX medium or into the periplasm respectively. High expression is

XX possible and the chemically induced overprod. is easy to handle.

XX Also, the prods. are homogeneous.

XX SAK-polypeptide derivs. are plasminogen activators for the

XX treatment of thrombosis.

alignment\_scores:  
 Quality: 695.00 Length: 136  
 Ratio: 5.148 Gaps: 0  
 Percent Similarity: 99.265 Percent Identity: 97.794

alignment\_block:  
 US-09-728-670-10 x AAQ44268 ..

Align seg 1/1 to: AAQ44268 from: 1 to: 414

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17
| | | | | | | | | | | | | | | | | | | | | | | |
4 TCAAGTTCATTGACAAAGAAATATATAAAAGCGATGACGCGATT 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
| | | | | | | | | | | | | | | | | | | | | | | |
54 TTTGAACCAACAGGCCCGTATTGATGTAATGTGACTGAGTTGATG 103
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
: | | | | | | | | | | | | | | | | | | | | | |
104 GTAAAGAAATGAATGCTATCCCTCGTTATGTCGAGTTTCCATTAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyr 67
| | | | | | | | | | | | | | | | | | | | | | | |
154 CCTGGACTACACTTACAAAGAAATTAATGATATGTCGATGGGC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
| | | | | | | | | | | | | | | | | | | | | | | |
204 ATTAGATGCGACAGCATATATAAGAGTTAGAGTAGTGAATTAGAT 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
| | | | | | | | | | | | | | | | | | | | | | | |
```



254 GCGCAAGATCGAGTCACCTTATATGATAGATAGAAAAAGAGAA 303  
101 ThrIysSerPheProIleThrGluLysGlyPheValProAspLeu 117  
|||||  
304 ACGAAGCTTTCCTATACAGAAAAAGGTTTGTGTGCCAGATTATC 353  
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134  
|||||  
354 AGAGCATATTAAAAACCTGGATTCACTTAATTACMAAGTTGTATAG 403  
134 LuLysLys 136  
|||||  
404 AAAAGAAA 411

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH49196

seq\_documentation\_block:  
ID AAH49196 standard; DNA; 540 BP.

AC AAH49196;

XX 26-NOV-2001 (first entry)

DE E. coli Lac-YH1-Sak construct DNA.

XX L-form bacteria production; recombinant protein production;  
KW temperature stress; lactose permease helix 1; ds.

OS Escherichia coli.  
OS Synthetic.

XX key Location/Qualifiers

FT misc\_structure 1..129

FT /tag= a

FT /note= "Helix 1 region"

XX MO200166776-A2.

XX 13-SEP-2001.

XX 08-MAR-2001; 2001WO-EP02630.

XX 10-MAR-2000; 2000DE-1011617.

PR 12-MAR-2000; 2000DE-1011358.

XX (MOLE-) INST MOLEKULARE BIOTECHNOLOGIE EV.

XX Hoischen C, Gumpert J, Kujau JM, Fritsche C, Eiske G, Fahnert B;

PI Sieben S, Mueller HP;

XX WPI; 2001-582277/65.

DR P-PSDB; AAB86832.

XX Producing a modified L-form bacterial strain by culturing the strain in

PT complex medium and subjecting it to temperature stress, provides

XX Claim 5; Fig 7a; 90pp; German.

XX This invention describes a novel method of producing a modified L-form  
CC bacterial strain which comprises culturing a bacterial strain adapted to  
CC a complex culture medium at alternate temperatures between 20 and 40  
CC plusoc, and fermenting the strain when hydromechanical stress of the  
CC cells increases. The modified L-form bacterial strains which are used to  
CC produce recombinant proteins. This sequence encodes the Escherichia  
CC coli derived construct LacYH1-Sak which is used to illustrate the method  
CC of the invention.

XX Sequence 540 BP; 192 A; 87 C; 99 G; 162 T; 0 other;

alignment\_scores:

Quality: 695.00

length: 136

Ratio: 5.148 Gaps: 0  
Percent Similarity: 99.265 Percent Identity: 97.794

alignment\_block:

US-09-728-670-10 x AAH49196 ..

Align seg 1/1 to: AAH49196 from: 1 to: 540

1 SerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTy 17  
|||||  
130 TCAAGTTCATTCGACAAAGAAAAATATATAAAAGCGATGACGCGAGTTA 179  
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
180 TTTTGACCAACAGCGCCGTATTGTGATGTAATGTGACTGAGATTGATG 229  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
::|||  
230 GTAAAGAAATGAATTGCTATCCCTCGTTATGTCGAGTTCTTATTA 279  
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67  
|||||  
280 CCTGGACTACACTTACAAAAGAAAAATGAATACTATGTCGAATGGGC 329  
67 aleuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspProS 84  
|||||  
330 ATTAGATGCGACAGCATATAAAGAGTTTGTAGTGAATTAGATCCAA 379  
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100  
|||||  
380 GCGCAAGATCGAGTCACCTTATATGATAGATAAGAAAAAGAGAA 429  
101 ThrIysSerPheProIleThrGluLysGlyPheValProAspLeu 117  
|||||  
430 ACGAAGTCTTCCCTATAACAGAAAAAGTTTGTGTGCCAGATTATC 479  
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134  
|||||  
480 AGAGCATATTAAAAACCTGGATTCACTTAATTACAAAGGTTGTATAG 529  
134 LuLysLys 136  
|||||  
530 AAAAGAAA 537

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ4276

seq\_documentation\_block:

ID AAQ4276 standard; DNA; 1023 BP.

AC AAQ4276;

XX 03-DEC-1993 (first entry)

DE Staphylokinase SAK42D plasmid pDB17.

XX Signal sequence, plasminogen activator; thrombosis; staphylokinase;

KW SAK; ss.

XX Staphylococcus aureus phage 42D.

XX key Location/Qualifiers

FT -35\_signal 224..229

FT /tag= a

FT -10\_signal 248..253

FT /tag= b

FT RBS 332..336

FT /tag= c

FT CDS 344..835

FT /tag= d

FT sig\_peptide 344..415

FT /tag= e

FT mat\_peptide 416..832

FT /tag= f

FT /label= SAK42D  
XX PN W09313209-A.  
XX PD 08-JUL-1993.  
XX PF 28-DEC-1992; 92WO-EP02989.  
XX PR 30-DEC-1991; 91DE-4143279.  
PR 22-JUN-1992; 92DE-4220516.  
PR 01-DEC-1992; 92DE-4240801.  
XX PA (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.  
XX PI Albrecht S, Behnke D, Guehrs K, Hartmann M, Schloft B;  
XX DR WPI: 1993-227325/28.  
DR P-PSDB; AAR39155.  
XX PT Staphylo-kinase (SAK) sequences lacking sequences for signal  
PT peptide(s) - for prodn. of proteins used as plasminogen  
PT activators in thrombosis treatment, and monoclonal antibodies  
PT against SAK  
XX PS Disclosure; Fig 9; 99pp; German.  
XX CC DNA encoding SAK lacking the signal peptide, is expressed  
CC intracellularly. This avoids the problem of fast degradation of the  
CC polypeptides or destruction of the host when expressed into the  
CC medium or into the periplasm respectively. High expression is  
CC possible and the chemically induced overprodn. is easy to handle.  
CC Also, the prods. are homogeneous.  
CC SAK-polypeptide derivs. are plasminogen activators for the  
CC treatment of thrombosis.  
XX SQ Sequence 1023 BP; 364 A; 123 C; 190 G; 346 T; 0 other;

alignment\_scores:                      Length: 136  
                    Quality: 695.00                      Gaps: 0  
                    Ratio: 5.148                      Percent Identity: 97.794  
Percent Similarity: 99.265

alignment\_block:  
US-09-728-670-10 x AAQ44276 ..

Align seg 1/1 to: AAQ44276 from: 1 to: 1023

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTyr 17  
|||||  
425 TCAAGTTCATTCGACAAAGAAATATATAAAAGGCGATGACGCGAGTTA 474

17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34  
|||||  
475 TTTTGAACCAACAGGCGCCGATTTGATGTTAAATGCTGAGTTGATG 524

34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
::|||  
525 GTAAAGAAATGAATGCTATCCCTCGTTATGTCGAGTTCCCTATTAA 574

51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrpA 67  
|||||  
575 CCGGGGACTACACTTACAAAGAAAAAATTGAATACTATGTCGAATGGGC 624

67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProS 84  
|||||  
625 ATTAGATGCGACAGCATATTAAGAGTTTAGAGTACTTGAATTAGATCCA 674

84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100  
|||||  
675 GCGCAAGATCGAAGTACTTATGATAGAAATAGAAAAAGAAAGAA 724

101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117

|||||  
725 ACGAAGCTTCCCTATAACAGAAAAAGTTTGTGTCGCCAGATTATTC 774

117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134  
|||||  
775 AGAGCATATTAAAAACCCCTGAGTTCACTTAATACAAAGCTTGTATAG 824

134 LuLysLys 136  
|||||  
825 AAAAGAAA 832

seq\_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAQ94334

seq\_documentation\_block:  
ID AAQ94334 standard; DNA; 492 BP.  
XX AC AAQ94334;  
XX DT 15-MAY-1996 (first entry)  
XX DE Recombinant staphylokinase gene SAK-1.  
XX KW Staphylokinase; Staphylococcus aureus; primer; PCR; amplification; lung;  
KW prokaryote; expression vector; E.coli; chromatography; thrombus; brain;  
KW peripheral blood vessel; heart attack; clogging; artery plug-in pipe;  
KW operation; eye; elution; ss.  
XX OS Staphylococcus aureus.  
XX FH Key Location/Qualifiers  
FH CDS 427..429  
FT /\*tag= a  
FT /transl\_except= seq: ATA, a.a.: leu

W09527048-A1.  
PN 12-OCT-1995.  
PD 03-APR-1995; 95WO-CN00025.  
PF 04-APR-1994; 94CN-0112105.  
PR (UYSH-) UNIV SHANGHAI MEDICAL.  
PA Song H;  
XX PI  
XX DR WPI: 1995-358627/46.  
DR P-PSDB; AAR75207.  
XX PT Prepn. of recombinant staphylokinase - used to cure thrombus caused  
PT by heart attack, blood clogging, etc.  
XX PS Examples; Page 11; 19pp; Chinese.  
XX CC The nucleotide sequence of the staphylokinase gene SAK-1 from  
CC Staphylococcus aureus. The gene is amplified by the primers AAQ94332-3  
CC and inserted into a prokaryotic expression vector such as pLY-4 to  
CC produce the plasmid pSTE-SAK-1. The plasmid is used to transform E.coli  
CC cells from where the kinase can be purified by a two-step  
CC chromatographic method. The kinase is used to treat thrombus in lungs,  
CC brains, and peripheral blood vessels after a heart attack, to prevent  
CC clogging in artery plug-in pipes during operations, blood accumulating in  
CC eyes and elution, etc.  
XX SQ Sequence 492 BP; 186 A; 68 C; 91 G; 147 T; 0 other;

alignment\_scores:                      Length: 136  
                    Quality: 693.00                      Gaps: 0  
                    Ratio: 5.133                      Percent Identity: 97.059  
Percent Similarity: 99.265

alignment\_block:

US-09-728-670-10 x AAQ94334 ..

Align seg 1/1 to: AAQ94334 from: 1 to: 492

```

1 SerSerSerPheAspLysGlyTyrLysLysGlyAspAlaSerTy 17
  |||
82 TCAAGTTCATTCGACAAAGAAATATATAAAGGCGATGACGAGTTA 131
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
  |||
132 TTTTGAACCAACAGCCCGCTATTTCATGCTAAATGTGACTGAGTTGATG 181
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
  ::|||
182 GTAAAGAAATGAATGCTATGCCCTCGTTATGTCGAGTTCCTATTAA 231
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
  |||
232 CCTGGGACTACACTTACAAAAGAAAATGAATACTATGTCGAATGGGC 281
67 aLeuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspPro 84
  |||
282 ATTAGATGCGACAGCATATTAAGAGTTTAGAGTAGTTGAATTAGATCCA 331
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
  |||
332 GCGCAAGATCGAAGTCACCTTATGATTAAGAATTAAGAAAAGAAAGAA 381
101 ThrLysSerPheProIleThrGlyLysGlyPheValProAspLeuSe 117
  |||
382 ACGAAGTCTTCCCTATTAACAGAAAAGTTTGTGTGCCAGATATATC 431
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134
  |||
432 AGAGCATATTAATAAACCCCTGATCACTTAATTACAAAGCTGTATAG 481
134 LuLysLys 136
  |||
482 AAAAGAAA 489

```

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44273

seq\_documentation\_block:

ID AAQ44273 standard; DNA; 414 BP.

```

XX AC AAQ44273;
XX DT 03-DEC-1993 (first entry)
XX DE Staphylokinase SAKM26C.
XX KW Signal sequence, plasminogen activator; thrombosis; staphylokinase;
XX KM SAK; ss.
XX OS Staphylococcus aureus phage 42D.
XX PN W09313209-A.
XX PD 08-JUL-1993.
XX PF 28-DEC-1992; 92WO-EP02989.
XX PR 30-DEC-1991; 91DE-4143279.
XX PR 22-JUN-1992; 92DE-4220516.
XX PR 01-DEC-1992; 92DE-4240801.
XX PA (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.
XX PI Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
XX DR WPI; 1993-227325/28.
XX DR P-PSDB; AAR39153.
XX

```

PT Staphylo-kinase (SAK) sequences lacking sequences for signal  
PT peptide(s) - for prodn. of proteins used as plasminogen  
PT activators in thrombosis treatment, and monoclonal antibodies  
PT against SAK  
XX  
PS Claim 5; Fig 6; 99pp; German.  
XX

CC DNA encoding SAK lacking the signal peptide, is expressed  
CC intracellularly. This avoids the problem of fast degradation of the  
CC polypeptides or destruction of the host when expressed into the  
CC medium or into the periplasm respectively. High expression is  
CC possible and the chemically induced overprod. is easy to handle.  
CC Also, the prods. are homogeneous.  
CC SAK-polypeptide derivs. are plasminogen activators for the  
CC treatment of thrombosis.  
XX

SO Sequence 414 BP; 158 A; 60 C; 81 G; 106 T; 9 other;

alignment\_scores:

Quality:	Ratio:	Length:
689.00	5.142	136
Percent Similarity: 98.529	Percent Identity: 97.059	Gaps: 0

alignment\_block:

US-09-728-670-10 x AAQ44273 ..

Align seg 1/1 to: AAQ44273 from: 1 to: 414

```

1 SerSerSerPheAspLysGlyTyrLysLysGlyAspAlaSerTy 17
  |||
4 TCAAGTTCATTCGACAAAGAAATATATAAAGGCGATGACGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
  |||
54 TTTGAACCAACAGCCCGTATTGTGTGATTAATGTGACTGAGATGAYC 103
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
  ::|||
104 GTAAAGAAATGARCTSYTRCCCKGKTATGTCGAGTTTCCCTATTAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
  |||
154 CCTGGGACTACACTTACAAAAGAAAATGAATACTATGTCGAATGGGC 203
67 aLeuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspPro 84
  |||
204 ATTAGATGCGACAGCATATTAAGAGTTTAGAGTAGTTGAATTAGATCCA 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
  |||
254 GCGCAAGATCGAAGTCACCTTATGATTAAGAAATTAAGAAAAGAAAGAA 303
101 ThrLysSerPheProIleThrGlyLysGlyPheValProAspLeuSe 117
  |||
304 ACGAAGTCTTCCCTATTAACAGAAAAGTTTGTGTGCCAGATTTATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134
  |||
354 AGAGCATATTAATAAACCCCTGATTCACACTTAATTACCAAGTTGTATAG 403
134 LuLysLys 136
  |||
404 AAAAGAAA 411

```

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44274

seq\_documentation\_block:

ID AAQ44274 standard; DNA; 414 BP.

```

XX AC AAQ44274;
XX DT 03-DEC-1993 (first entry)
XX

```

```
XX Staphylokinase SAKM26L.
DE
XX Signal sequence, plasminogen activator; thrombosis; staphylokinase;
KW SAK; ss.
OS Staphylococcus aureus phage 42D.
XX
XX WO9313209-A.
XX
XX 08-JUL-1993.
XX
XX 28-DEC-1992; 92WO-EP02989.
XX
XX 30-DEC-1991; 91DE-4143279.
XX 22-JUN-1992; 92DE-4220516.
XX 01-DEC-1992; 92DE-4240801.
XX
XX (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.
XX
XX Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
PI
XX WPI; 1993-227325/28.
XX
XX P-PSDB; AAR39154.
XX
XX Staphylo-kinase (SAK) sequences lacking sequences for signal
PT peptide(s) - for prodn. of proteins used as plasminogen
PT activators in thrombosis treatment, and monoclonal antibodies
PT against SAK
XX
XX Claim 5; Fig 7; 99pp; German.
XX
XX DNA encoding SAK lacking the signal peptide, is expressed
CC intracellularly. This avoids the problem of fast degradation of the
CC polypeptides or destruction of the host when expressed into the
CC medium or into the periplasm respectively. High expression is
CC possible and the chemically induced overprodn. is easy to handle.
CC Also, the prods. are homogeneous.
CC SAK-polypeptide derivs. are plasminogen activators for the
CC treatment of thrombosis.
XX
XX Sequence 414 BP; 159 A; 60 C; 81 G; 105 T; 9 other;
SQ

alignment_scores:
Quality: 684.00 Length: 136
Ratio: 5.143 Gaps: 0
Percent Similarity: 97.794 Percent Identity: 96.324

alignment_block:
US-09-728-670-10 x AAQ44274 ..

Align seg 1/1 to: AAQ44274 from: 1 to: 414

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
4 TCAAGTTCATTCGACAAAGGAAATATATAAAAGGCGATGACGCGAGTTA 53
|||||
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsps 34
|||||
54 TTTTGAACACACAGCGCCCTATTGTGACGCTAAATGTGACTGAGYTGAYG 103
|||||
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
::|||
104 GTAAAGAAATGARCTSTRTCCCKCGKTATGTCGAGTTCTTAA 153
|||||
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrpAl 67
|||||
154 CCGGCGACTACACTTACAAAGAAATGAATGACTATGTCGAATGGGC 203
|||||
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
204 ATTAGATGCGACACGACATATAAAGAGTTTAGAGTAGTTGAATTAGATCCA 253
|||||
```

```
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
254 GCGCAAGATCGAAGTCACCTATTATGATAGAAATAAGAAAGAGAA 303
|||||
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
|||||
304 ACGAAGTCTTCCCTATACAGAAAAAGGTTTGTGTCGACAGATTATC 353
|||||
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
134 LuLysLys 136
|||||
404 AAAAGAAA 411
|||||

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:AAAN70606
seq_documentation_block:
ID AAAN70606 standard; DNA; 991 BP.
XX
XX AAAN70606;
XX
XX 18-FEB-1991 (first entry)
XX
XX Sequence of staphylokinase gene from Staphylococcus aureus phage of
DE serological gp. F esp. the phase phi 42D.
XX
XX Enzyme; thrombolytic agent; bacterium; plasminogen; plasmin; ss.
XX
XX Staphylococcus aureas phage gp. F.
XX
XX Key Location/Qualifiers
FH CDS 314..805
FT -35_signal /*tag= a 194..199
FT -10_signal /*tag= b 218..223
FT RBS /*tag= c 302..307
FT /*tag= d

DD245444-A.
PN
XX 06-MAY-1987.
PD
XX 31-JAN-1986; 86DD-0286669.
PF
XX 31-JAN-1986; 86DD-0286669.
PR
XX 31-JAN-1986; 86DD-0286669.
XX
XX (DEAK ) AKAD WISSENSCHAFT DDR.
XX
XX Behnke D, Gerlach D, Adler B;
PI
XX WPI; 1987-264521/38.
DR
XX P-PSDB; AAP70381.
XX
XX New staphylokinase prodn. from heterologous prodn. cells -
PT transformed with recombinant plasmids. useful as thrombolytic
PT agent
XX
XX Disclosure; Fig 1, Page 10; 16pp; German.
XX
XX AAAN70606 is the SQ of the pref. gene for a new method of prodn. of
CC staphylokinase. Using the new method, staphylokinase can be prepd.
CC uncontaminated by toxic or antigenic materials and where Gram
CC positive prodn. strains are used, it is excreted into the culture
CC medium, facilitating recovery.
XX
XX Sequence 991 BP; 356 A; 118 C; 183 G; 334 T; 0 other;
SQ
```



## alignment\_scores:

Quality: 672.00 Length: 136  
 Ratio: 5.053 Gaps: 0  
 Percent Similarity: 97.794 Percent Identity: 94.853

## alignment\_block:

US-09-728-670-10 x AAN70606 ..

Align seg 1/1 to: AAN70606 from: 1 to: 991

```

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
  |||||||
395 TCAAGTTCATTCGACAAAGAAATATATAAAAAAGCGATGACCGAGTTA 444
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
  |||||||
445 TTTTGAACCAACAGCGCCGTTTGTGATGATGATGATGATGATGATG 494
34 eLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
  :|||
495 GTAAAGAAATGAATTCCTATCCCTCGTATGTCGAGTTTCTATTAA 544
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTTPAl 67
  |||||||
545 CCTGGGACTACACTTACAAAGAAATGAATACTATCTCGAATGGGC 594
67 aLeuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspProS 84
  |||||||
595 ATTAGATGCGACAGATATAAAGATTAGAGTAGTGAATTAGATCCAA 644
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
  |||||||
645 GCGCAAGATCGAAGTCACCTATATAGATAAGATAAGAAAGAAAGAA 694
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
  |||||||
695 ACGAGTCTTCCCTATACAGAAAGTTTGTGTGTCAGATTATC 744
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
  |||||||
745 AGAGCATATTAAAAACCTGATTCACCTTAATTACAAAGCTGTATTAG 794
134 LuLysLys 136
  |||||||
795 AAAAGAAA 802

```

seq\_name: /SIDSl/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44271

## seq\_documentation\_block:

ID AAQ44271 standard; DNA; 384 BP.

```

XX AAQ44271;
AC AAQ44271;
DT 03-DEC-1993 (first entry)
XX
DE Staphylokinase SAKdeltaN10.
XX
KW Signal sequence, plasminogen activator; thrombosis; staphylokinase;
  SAK; ss.
XX
OS Staphylococcus aureus phage 42D.
XX
PN WO9313209-A.
XX
PD 08-JUL-1993.
XX
PF 28-DEC-1992; 92WO-EP02989.
XX
PR 30-DEC-1991; 91DE-4143279.
PR 22-JUN-1992; 92DE-4220516.
PR 01-DEC-1992; 92DE-4240801.
XX
PA (MEDA-) MEDAC GES KLINISCHE SPEZIALPRAEPARATE.

```

```

XX Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
XX WPI; 1993-227325/28.
DR P-PSDB; AAR39151.
XX
PT Staphylo-kinase (SAK) sequences lacking sequences for signal
PT peptide(s) - for prodn. of proteins used as plasminogen
PT activators in thrombosis treatment, and monoclonal antibodies
PT against SAK
XX
PS Claim 4; Fig 4; 99pp; German.
XX
CC DNA encoding SAK lacking the signal peptide, is expressed
CC intracellularly. This avoids the problem of fast degradation of the
CC polypeptides or destruction of the host when expressed into the
CC medium or into the periplasm respectively. High expression is
CC possible and the chemically induced overprod. is easy to handle.
CC Also, the prods. are homogeneous.
CC SAK-polypeptide derivs. are plasminogen activators for the
CC treatment of thrombosis.
XX
SQ Sequence 384 BP; 146 A; 55 C; 78 G; 104 T; 1 other;

```

## alignment\_scores:

Quality: 643.00 Length: 126  
 Ratio: 5.144 Gaps: 0  
 Percent Similarity: 99.206 Percent Identity: 97.619

## alignment\_block:

US-09-728-670-10 x AAQ44271 ..

Align seg 1/1 to: AAQ44271 from: 1 to: 384

```

11 LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVa 27
  |||||||
4 AAAAGCGATGACGCGAGTTATTTGAACCAACAGCGCCGTTATTTGATG 53
27 lAsnValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHisT 44
  |||||||
54 AAATGTGACTGAGTTGATGTGTAAGAAGAAATGCTATCCCTCGTT 103
44 yRValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIle 60
  |||||||
104 ATGTCGAGTTTCCCTATTAAACCTGGGACTACACTTACAAAGAAATTT 153
61 GluTyrTyrValGluTTPAlaLeuAspAlaThrAlaTyrLysGluPheAr 77
  |||||||
154 GAATACTATGTGAAATGGGCATTAGATGCGACGACATATAAAGAGTTAG 203
77 gValValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspL 94
  |||||||
204 AGTAGTTGAATTGATTCCAAGCGCAAGATCGAAGTCACCTATTATGATA 253
94 yAsnLysLysLysGluGluThrLysSerPheProIleThrGluLysGly 110
  |||||||
254 AGAATTAAGAAAAAGAAAGAAAGCAAGTCTTCCCTATTAACAGAAAAAG 303
111 PheValValProAspLeuSerGluHisIleLysAsnProGlyPheAsnLe 127
  |||||||
304 TTGCTGTGCCAGATTATCAGAGCATATTAAAAACCTGATTCACCTT 353
127 uIleThrLysValValIleGluLysLys 136
  |||||||
354 AATTACMAAGGTTGTTATAGAAAAAGAA 381

```

seq\_name: /SIDSl/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ44272

## seq\_documentation\_block:

ID AAQ44272 standard; DNA; 372 BP.

XX

AC AAQ44272;

```
XX 03-DEC-1993 (first entry)
DT Staphylokinase SAKdeltaN14.
DE
XX Signal sequence, plasminogen activator; thrombosis; staphylokinase;
KW SAK; ss.
XX
OS Staphylococcus aureus phage 42D.
XX
PN W09313209-A.
XX
PD 08-JUL-1993.
XX
PF 28-DEC-1992; 92WO-EP02989.
XX
PR 30-DEC-1991; 91DE-4143279.
PR 22-JUN-1992; 92DE-4220516.
PR 01-DEC-1992; 92DE-4240801.
XX
PA (MEDA-) MEDAC GES KLINISCHE SPEZIALPREPARATE.
XX
PI Albrecht S, Behnke D, Guehrs K, Hartmann M, Schlott B;
XX
DR WPI; 1993-227325/28.
DR P-PSDB; AAR39152.
XX
XX Staphylo-kinase (SAK) sequences lacking sequences for signal
PT peptide(s) - for prodn. of proteins used as plasminogen
PT activators in thrombosis treatment, and monoclonal antibodies
PT against SAK
XX
PS Claim 4; Fig 5; 99pp; German.
XX
CC DNA encoding SAK lacking the signal peptide, is expressed
CC intracellularly. This avoids the problem of fast degradation of the
CC polypeptides or destruction of the host when expressed into the
CC medium or into the periplasm respectively. High expression is
CC possible and the chemically induced overprodn. is easy to handle.
CC Also, the prods. are homogeneous.
CC SAK-polypeptide derivs. are plasminogen activators for the
CC treatment of thrombosis.
XX
SQ Sequence 372 BP; 141 A; 53 C; 74 G; 103 T; 1 other;

alignment_scores:
Quality: 620.00 Length: 122
Ratio: 5.124 Gaps: 0
Percent Similarity: 99.180 Percent Identity: 97.541

alignment_block:
US-09-728-670-10 x AAQ44272 ..

Align seg 1/1 to: AAQ44272 from: 1 to: 372

15 AlasERTyrPheGluProThrglyProTyrLeuMetValAsnValThrgl 31
|||||
4 GCGAGTTATTGAAACCAACAGGCCCGTATTGATGTAATGTGACTGG 53
31 yValAspSerLysGlyAsnGluLeuLeuSerProHisTyrValGluPhep 48
|||||
54 AGTTGATGTAAAGAAATGAATGCTATCCCTCGTATGTCGAGTTTC 103
48 roileLysProGlyThrThrLeuThrLysGluLysIleGluTyrTyrVal 64
|||||
104 CTATTAAACCTGGGACTACACTTACAAAGAAATAATGAATACTATGTC 153
65 GluTrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValValGluLe 81
|||||
154 GAATGGGCATTAGATGCACAGCATATAAGAGTTTGAAGTAGTGAATT 203
81 uAspProSerAlaLysIleGluValThrTyrTyrAspLysAsnLysLysL 98
```

```
|||||
204 AGATCCAAAGCGCAAGATCGAGTCACCTTATTATGATTAAGAATAAGAAAA 253
98 ysgLgluThrLysSerPheProIleThrgluLysGlyPheValValPro 114
|||||
254 AAGAAGAAACGAAGTCTTCCCTATACAGAAAAAGGTTTGTGTCCCA 303
115 AspleuSerGluHisIleLysAsnProGlyPheAsnLeuIleThrlsVa 131
|||||
304 GATTATTCAGAGCATATTAAACCCCTGATTCACCTTAATTACMAAGGT 353
131 lValIleGluLysLys 136
|||||
354 TGTATTAGAAAAAGAAA 369
```

seq\_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:AA60717

seq\_documentation\_block:

ID AA60717 standard; DNA; 634 BP.

XX AA60717;

DT 23-OCT-1991 (first entry)

DE Fragment of sak gene in novel compound plasmid.

KW Colibacillus; sak promoter; psak; SIGsak; ss.

OS Staphylococcus aureus.

FH Key Location/Qualifiers

FT promoter 191..220

FT RBS 299..303

FT CDS 311..634

FT sig-peptide 311..391

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

FT /\*tag= d

JP61135590-A.

23-JUN-1986.

05-DEC-1984; 84JP-0257258.

05-DEC-1984; 84JP-0257258.

(HONS ) YAKULT HONSHA KK.

WPI; 1986-267216/41.

P-PSDB; AAP60796.

Compound plasmid - obtd. by joining three specified DNA pieces to

colon bacterium plasmid vector.

Disclosure; Fig 5; 17pp; Japanese.

The sequence is part of a compound plasmid construct comprising a

colibacillus promoter fragment, a S.aureus sak gene promoter (Psak)

ribosome binding site and signal peptide (SIGsak) and a restriction

enzyme recognition site at which an heterogene is inserted.

alignment\_scores:

Quality: 417.00 Length: 81

Ratio: 5.213 Gaps: 0

Percent Similarity: 98.765 Percent Identity: 97.531

```
alignment_block:
US-09-728-670-10 x AAN60717 ..
Align seg 1/1 to: AAN60717 from: 1 to: 634

1 SerSerSerPheaspIysGlyLysTyrLysGlyAspAlaSerTy 17
|||||
392 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACCGAGTTA 441
17 rPhegluProthrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
442 TTTTGAACCAACAGCGCGGTATTGATGGTAAATGTGACTGGAGTTGATG 491
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::
492 GTAAAGGAATGAATTGCTATCCCTCATTTATGTGAGTTCTCTATTAA 541
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrpAl 67
|||||
542 CTTGGGACTACACTTACAAAGAAAAAATTTGAATTAATGTCGAATGGGC 591
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeu 81
|||||
592 ATTAGATGCGACAGCAATTAAGAGTTTAGAGTAGTGAATTGA 634

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:AAQ24798
seq_documentation_block:
ID AAQ24798 standard; DNA; 490 BP.
XX AAQ24798;
XX 17-NOV-1992 (first entry)
DE SAK/trypstatin fusion protein.
XX
KW Staphylokinase; alkaline phosphatase; signal peptide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 52..483
FT /tag= a
FT /label= SAK-trypstatin_fusion_protein
FT misc_feature 52..288
FT /tag= b
FT misc_feature 289..300
FT /label= SAK_coding_region
FT /tag= c
FT /label= Linker
FT misc_feature 301..483
FT /tag= d
FT /label= Trypstatin_structural_gene
XX
PN JP04112791-A.
XX
PD 14-APR-1992.
XX
PE 31-AUG-1990; 90JP-0232268.
XX
PR 31-AUG-1990; 90JP-0232268.
XX
PA (TAIS ) TAISHO PHARM CO LTD.
XX
DR WPI; 1992-173147/21.
XX
PT Trypstatin structural gene with specified DNA base sequence -
PT used with specified DNA to form expression vector, for commercial
PT prodn. of trypstatin
XX
PS Disclosure; Fig 5; 18pp; Japanese.
XX
XX The sequences in AAQ24797-8 show two fusion constructs of SAK and a
```

```
CC synthetic trypstatin structural gene. This sequence may be linked
CC to a DNA sequence encoding the signal peptide of the staphylokinase
CC gene originated from Staphylococcus aureus or of the alkaline
CC phosphatase gene originated from E.coli. This expression vector can
CC then be transformed into E. coli or Bacillus subtilis and the SAK-
CC trypstatin fusion protein will be secreted extracellularly or in
CC the periplasm. It can then be recovered on a commercial scale.
XX
SQ Sequence 490 BP; 151 A; 90 C; 107 G; 142 T; 0 other;

alignment_scores:
Quality: 277.00 Length: 52
Ratio: 5.327 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.077

alignment_block:
US-09-728-670-10 x AAQ24798 ..
Align seg 1/1 to: AAQ24798 from: 1 to: 490

1 SerSerSerPheaspIysGlyLysTyrLysGlyAspAlaSerTy 17
|||||
133 TCAAGTTCATTCGACAAAGAAATTAATAAAAGCGATGACCGAGTTA 182
17 rPhegluProthrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
183 TTTTGAACCAACAGCGCGGTATTGATGGTAAATGTGACTGGAGTTGATG 232
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::
233 GTAAAGGAATGAATTGCTATCCCTCATTTATGTGAGTTCTCTATTAA 282
51 ProGly 52
|||||
283 CTTGGG 288

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:AAQ11814
seq_documentation_block:
ID AAQ11814 standard; DNA; 480 BP.
XX AAQ11814;
XX 05-AUG-1991 (first entry)
DE Encodes staphylokinase-somatomedin C fusion protein.
XX
KW Staphylococcus aureus; SAK; SMC; protein production; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT sig_peptide 1..81
FT /tag= a
FT mat_peptide 262..474
FT /tag= b
FT misc_RNA 237..261
FT /product= somatomedin C
FT /tag= c
FT mat_peptide 82..236
FT /function= linker
FT /tag= d
FT /product= Staphylokinase N-terminal region
XX
PN JP03098595-A.
XX
PD 24-APR-1991.
XX
PF 11-SEP-1989; 89JP-0234874.
XX
PR 11-SEP-1989; 89JP-0234874.
XX
```

PA (TAIS ) TAISHO PHARMACEUT KK.  
XX  
DR WPI; 1991-167039/23.  
DR P-PSDB; AAR12138.  
XX  
PT prepn. of peptide(s) - by construction of expression  
PT vector, transformation of E.coli etc., culturing to secrete  
PT peptide(s) and collecting peptide(s)  
XX  
PS Example; Fig 4; 15pp; Japanese.  
XX  
CC This sequence encodes a SAK-SMC fusion protein comprising the  
CC Staphylokinase signal peptide and part of the SAK mature protein  
CC fused (via a linker) to somatomedin C. Expression of the fusion  
CC protein in recombinant hosts is under control of staphylococcal  
CC regulatory sequences.  
CC See also AAQ11813.  
XX  
SQ Sequence 480 BP; 125 A; 102 C; 107 G; 146 T; 0 other;

alignment\_scores:  
Quality: 271.00 Length: 51  
Ratio: 5.314 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.039

alignment\_block:  
US-09-728-670-10 x AAQ11814 ..

Align seg 1/1 to: AAQ11814 from: 1 to: 480

```
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
82 TCAAGTTCATTCGACAAAGAAATATATAAAGGCGATGACGCGAGTTA 131
17 rphegluprothrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
132 TTTTGAACCAACAGCGCCGTAATTGATGCTAAATGTGACTGCGAGTTGATG 181
34 erlysglyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::
182 GTAAAGGAATGAATTGCTATCCCTCATTTATGTCGAGTTTCCTATTAAA 231
```

51 Pro 51  
|||  
232 CCT 234

seq\_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:AAQ24797

seq\_documentation\_block:

ID AAQ24797 standard; DNA; 495 BP.

XX AAQ24797;

DT 17-NOV-1992 (first entry)

XX SAK/tryptstatin fusion protein.

XX Staphylokinase; alkaline phosphatase; signal peptide; ds.

OS Synthetic.

XX Key Location/Qualifiers

FH CDS 60..485

FT /\*tag= a

FT /label= SAK-tryptstatin\_fusion\_protein

FT misc\_feature 60..296

FT /\*tag= b

FT /label= SAK\_coding\_region

FT misc\_feature 297..302

FT /\*tag= c

FT /label= Linker

FT misc\_feature 303..485

FT /\*tag= d  
FT /label= tryptstatin\_structural\_gene

XX JP04112791-A.

XX 14-APR-1992.

XX 31-AUG-1990; 90JP-0232268.

XX 31-AUG-1990; 90JP-0232268.

XX (TAIS ) TAISHO PHARM CO LTD.

XX WPI; 1992-173147/21.

XX Tryptstatin structural gene with specified DNA base sequence -

XX used with specified DNA to form expression vector, for commercial

XX prodn. of tryptstatin

XX Disclosure; Fig 5; 18pp; Japanese.

XX The sequences in AAQ24797-8 show two fusion constructs of SAK and a  
CC synthetic tryptstatin structural gene. This sequence may be linked  
CC to a DNA sequence encoding the signal peptide of the staphylokinase  
CC gene originated from Staphylococcus aureus or of the alkaline  
CC phosphatase gene originated from E.coli. This expression vector can  
CC then be transformed into E. coli or Bacillus subtilis and the SAK-  
CC tryptstatin fusion protein will be secreted extracellularly or in  
CC the periplasm. It can then be recovered on a commercial scale.

SQ Sequence 495 BP; 152 A; 91 C; 106 G; 146 T; 0 other;

alignment\_scores:  
Quality: 264.00 Length: 53  
Ratio: 5.077 Gaps: 1  
Percent Similarity: 98.113 Percent Identity: 96.226

alignment\_block:  
US-09-728-670-10 x AAQ24797 ..

Align seg 1/1 to: AAQ24797 from: 1 to: 495

```
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
141 TCAAGTTCATTCGACAAAGAAATATATAAAGGCGATGACGCGAGTTA 190
17 rphegluprothrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
191 TTTTGAACCAACAGCGCCGTAATTGATGCTAAATGTGACTGCGAGTTGATG 240
34 erlysglyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:::
241 GTAAAGGAATGAATTGCTATCCCTCATTTATGTCGAGTTTCCTATTAA 290
50 SPROGly 52
|||||
291 ACCTGGG 297
```

seq\_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAC85121

seq\_documentation\_block:

ID AAC85121 standard; DNA; 66 BP.

XX AAC85121;

DT 08-MAY-2001 (first entry)

XX S. aureus staphylokinase mutagenesis primer 2d (wt).

XX Immunogenicity; staphylokinase; variant; stability; mutagenesis;

KW PCR primer; ss.

XX



```
OS Staphylococcus aureus.
XX
PN WO200104287-A1.
XX
PD 18-JAN-2001.
XX
PF 06-JUL-2000; 2000WO-DK00371.
XX
PR 07-JUL-1999; 99DK-0000988.
PR 27-AUG-1999; 99DK-0001196.
PR 02-MAR-2000; 2000DK-0000339.
PR 18-MAY-2000; 2000DK-0000804.
XX
PA (MAXY-) MAXYGEN APS.
XX
PI Halkier T, Pedersen AH, Okkels JS;
XX
DR WPI; 2001-138342/14.
XX
PT Producing polypeptides with altered immunogenicity or improved
PT stability, comprises expressing a diversified nucleotide sequence
PT population and selecting polypeptides with altered immunogenicity or
PT improved stability -
XX
PS Example 2; Page 62; 83pp; English.
XX
CC The invention relates to a method of altering immunogenicity and/or
CC increasing stability of a polypeptide of interest. The method comprises
CC (a) expressing a diversified population of nucleotide sequences encoding
CC a polypeptide of interest; (b) screening the polypeptides expressed for
CC function, immunogenicity and/or stability; and (c) selecting functional
CC polypeptides with altered immunogenicity and/or increased stability. The
CC method is used to improve the properties of polypeptides, in particular
CC to alter the immunogenicity and/or increase the functional in vivo half-
CC life of the polypeptide. The method uses a high throughput system that
CC makes it possible to search several orders of magnitude more polypeptides
CC than is possible by previously known approaches. This enhances the chance
CC of finding the optimal variant from the many thousands of variants that
CC may be produced. Sequences AAC85113-121 represent PCR primers used in a
CC focused mutagenesis reaction towards introduction of lysine residues in a
CC S. aureus staphylokinase.
XX
SQ Sequence 66 BP; 18 A; 10 C; 9 G; 29 T; 0 other;

alignment_scores:
    Quality: 93.00      Length: 18
    Ratio: 5.167      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-728-670-10 x AAC85121/rev ..
Align seg 1/1 to reverse of: AAC85121 from: 1 to: 66

119 HisilelysasnpGlyPheasnleuilerHrlysvaIvalilegluly 135
|||||
65 CATATATTAACCAACCTGATTCACCTTAATTACAAAGTTGTATAGAAAA 16

135 slys 136
||||
15 GAAA 12

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAC85122
seq_documentation_block:
ID AAC85122 standard; DNA; 69 BP.
XX
AC AAC85122;
XX
DT 08-MAY-2001 (first entry)
XX
DE S. aureus staphylokinase mutagenesis primer 3.
```

```
XX
KW Immunogenicity; staphylokinase; variant; stability; mutagenesis;
KW PCR primer; ss.
XX
OS Staphylococcus aureus.
XX
PN WO200104287-A1.
XX
PD 18-JAN-2001.
XX
PF 06-JUL-2000; 2000WO-DK00371.
XX
PR 07-JUL-1999; 99DK-0000988.
PR 27-AUG-1999; 99DK-0001196.
PR 02-MAR-2000; 2000DK-0000339.
PR 18-MAY-2000; 2000DK-0000804.
XX
PA (MAXY-) MAXYGEN APS.
XX
PI Halkier T, Pedersen AH, Okkels JS;
XX
DR WPI; 2001-138342/14.
XX
PT Producing polypeptides with altered immunogenicity or improved
PT stability, comprises expressing a diversified nucleotide sequence
PT population and selecting polypeptides with altered immunogenicity or
PT improved stability -
XX
PS Example 3; Page 64; 83pp; English.
XX
CC The invention relates to a method of altering immunogenicity and/or
CC increasing stability of a polypeptide of interest. The method comprises
CC (a) expressing a diversified population of nucleotide sequences encoding
CC a polypeptide of interest; (b) screening the polypeptides expressed for
CC function, immunogenicity and/or stability; and (c) selecting functional
CC polypeptides with altered immunogenicity and/or increased stability. The
CC method is used to improve the properties of polypeptides, in particular
CC to alter the immunogenicity and/or increase the functional in vivo half-
CC life of the polypeptide. The method uses a high throughput system that
CC makes it possible to search several orders of magnitude more polypeptides
CC than is possible by previously known approaches. This enhances the chance
CC of finding the optimal variant from the many thousands of variants that
CC may be produced. Sequences AAC85122-123 represent PCR primers used in a
CC localised mutagenesis reaction to remove amino acid residues containing
CC attachment groups from a S. aureus staphylokinase.
XX
SQ Sequence 69 BP; 30 A; 13 C; 10 G; 16 T; 0 other;

alignment_scores:
    Quality: 92.00      Length: 27
    Ratio: 4.000      Gaps: 2
    Percent Similarity: 85.185    Percent Identity: 81.481

alignment_block:
US-09-728-670-10 x AAC85122 ..
Align seg 1/1 to: AAC85122 from: 1 to: 69

83 ProSerAlaIysIleGluValThrTyrTyraPlysaAsnLysLysG1 99
|||||
1 CCAAGCGCTAAGATCGAAGTCACCTATATATGAT.....AATAAGA 41

99 uGluThrLysSerPheProIleThrGluLys 109
|||||
42 AGAAACG...TCCTTCCCTATACAGAAAAA 69

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAC85119
seq_documentation_block:
ID AAC85119 standard; DNA; 66 BP.
XX
AC AAC85119;
```

```
XX 08-MAY-2001 (first entry)
DT
XX
DE S. aureus staphylokinase mutagenesis primer 2c.
XX
XX Immunogenicity; staphylokinase; variant; stability; mutagenesis;
KW PCR primer; ss.
XX
OS Staphylococcus aureus.
XX
XX WO200104287-A1.
XX
XX 18-JAN-2001.
XX
XX 06-JUL-2000; 2000WO-DK00371.
XX
XX 07-JUL-1999; 99DK-0000988.
XX
XX 27-AUG-1999; 99DK-0001196.
XX
XX 02-MAR-2000; 2000DK-0000339.
XX
XX 18-MAY-2000; 2000DK-0000804.
XX
XX (MAXY-) MAXYGEN APS.
XX
XX Halkier T, Pedersen AH, Okkels JS;
XX
XX WPI; 2001-138342/14.
XX
XX Producing polypeptides with altered immunogenicity or improved
PT stability, comprises expressing a diversified nucleotide sequence
PT population and selecting polypeptides with altered immunogenicity or
PT improved stability -
XX
XX Example 2; Page 62; 83pp; English.
XX
XX The invention relates to a method of altering immunogenicity and/or
CC increasing stability of a polypeptide of interest. The method comprises
CC (a) expressing a diversified population of nucleotide sequences encoding
CC a polypeptide of interest; (b) screening the polypeptides expressed for
CC function, immunogenicity and/or stability; and (c) selecting functional
CC polypeptides with altered immunogenicity and/or increased stability. The
CC method is used to improve the properties of polypeptides, in particular
CC to alter the immunogenicity and/or increase the functional in vivo half-
CC life of the polypeptide. The method uses a high throughput system that
CC makes it possible to search several orders of magnitude more polypeptides
CC than is possible by previously known approaches. This enhances the chance
CC of finding the optimal variant from the many thousands of variants that
CC may be produced. Sequences AAC85113-121 represent PCR primers used in a
CC focused mutagenesis reaction towards introduction of lysine residues in
CC S. aureus staphylokinase.
XX
XX Sequence 66 BP; 18 A; 11 C; 8 G; 29 T; 0 other;
SQ

alignment_scores:
    Quality: 87.00      Length: 18
    Ratio: 4.833        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 94.444

alignment_block:
US-09-728-670-10 x AAC85119/rev ..
Align seg 1/1 to reverse of: AAC85119 from: 1 to: 66

119 HisIleLysAsnProGlyPheAsnLeuIleThrLysValIleGluLys 135
|||||
65 CATATTAAAAACCCCTGATTCAGTTAATTACAAAGGTTGTATAGAAAA 16

135 slys 136
|||||
15 GAAA 12

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAC85120
```

```
seq_documentation_block:
ID AAC85120 standard; DNA; 66 BP.
XX
XX AAC85120;
AC
XX
XX 08-MAY-2001 (first entry)
DT
XX
XX S. aureus staphylokinase mutagenesis primer 2d.
DE
XX
XX Immunogenicity; staphylokinase; variant; stability; mutagenesis;
KW PCR primer; ss.
XX
XX Staphylococcus aureus.
XX
XX WO200104287-A1.
XX
XX 18-JAN-2001.
XX
XX 06-JUL-2000; 2000WO-DK00371.
XX
XX 07-JUL-1999; 99DK-0000988.
XX
XX 27-AUG-1999; 99DK-0001196.
XX
XX 02-MAR-2000; 2000DK-0000339.
XX
XX 18-MAY-2000; 2000DK-0000804.
XX
XX (MAXY-) MAXYGEN APS.
XX
XX Halkier T, Pedersen AH, Okkels JS;
XX
XX WPI; 2001-138342/14.
XX
XX Producing polypeptides with altered immunogenicity or improved
PT stability, comprises expressing a diversified nucleotide sequence
PT population and selecting polypeptides with altered immunogenicity or
PT improved stability -
XX
XX Example 2; Page 62; 83pp; English.
XX
XX The invention relates to a method of altering immunogenicity and/or
CC increasing stability of a polypeptide of interest. The method comprises
CC (a) expressing a diversified population of nucleotide sequences encoding
CC a polypeptide of interest; (b) screening the polypeptides expressed for
CC function, immunogenicity and/or stability; and (c) selecting functional
CC polypeptides with altered immunogenicity and/or increased stability. The
CC method is used to improve the properties of polypeptides, in particular
CC to alter the immunogenicity and/or increase the functional in vivo half-
CC life of the polypeptide. The method uses a high throughput system that
CC makes it possible to search several orders of magnitude more polypeptides
CC than is possible by previously known approaches. This enhances the chance
CC of finding the optimal variant from the many thousands of variants that
CC may be produced. Sequences AAC85113-121 represent PCR primers used in a
CC focused mutagenesis reaction towards introduction of lysine residues in
CC S. aureus staphylokinase.
XX
XX Sequence 66 BP; 16 A; 11 C; 9 G; 30 T; 0 other;
SQ

alignment_scores:
    Quality: 86.00      Length: 18
    Ratio: 5.059        Gaps: 0
    Percent Similarity: 94.444    Percent Identity: 94.444

alignment_block:
US-09-728-670-10 x AAC85120/rev ..
Align seg 1/1 to reverse of: AAC85120 from: 1 to: 66

119 HisIleLysAsnProGlyPheAsnLeuIleThrLysValIleGluLys 135
|||||
65 CATATTAAAAACCCCTGATTCAGTTAAGACAAAGGTTGTATAGAAAA 16

135 slys 136
|||||
```



alignment\_block:  
US-09-728-670-10 x AAC91695/rev ..  
Align seg 1/1 to reverse of: AAC91695 from: 1 to: 11225  
43 HistyrValGluPheProIleLeysProGlyThrThrLeuThrLysGlu 59  
||||| ||||| ::::::: ||::::: ||  
11113 CATATGTGTTGAACCTATACATATAGCTGCTAATAGAAAGAAAGTAA 11064  
59 sile...GluTyrTyrValGluTrpAlaLeuAspAlaThrAla..... 72  
||||| ::::: ||| ::::: ||  
11063 AATAGTAAATGCTTATAGATACGCGCTGACATTAATTCAGAGACG 11014  
73 .....TyrLysGlu 75  
11013 GCGCGAATGTAATACCTATACATTACGCTATGAAGTATACGATCCG 10964  
76 PheArgVal.....ValGluLeuAspProSerAlaLysIleG1 88  
||||| ::::: ||| ::::: ||  
10963 TTTAGATTAAATTAATTAAGTATTATTAGACCACGCGCC..... 10922  
88 uValThrTyrTyrAspLysAsnLysLysGluGluThrLysSerPheP 105  
||||| ::::: ||| ::::: ||  
10921 .....GATATTACAAACAAAGCGTTTAACTAATACATATCC 10885  
105 roIeThrGluLysGlyPheValValProAspLeuSerGluHisIleLys 121  
||||| ::::: ||| ::::: ||  
10884 CCTTATACGAAACTAGGTTTATACCGACGACCTATTAGATTACATCATA 10835  
122 AsnProGlyPheAsnLeu.....IleTh 129  
::: ||| ::::: |||  
10834 TCTAGAGAGAGCTAATATATAATAAAGAAAGATGGGTAGATATATTAC 10785  
129 rLysValValIleGluLys 135  
::: ||| ::::: |||  
10784 ACGAATAATATTAGAACGA 10766  
seq\_name: /SIDSl/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAC85118  
seq\_documentation\_block:  
ID AAC85118 standard; DNA; 66 BP.  
XX AAC85118;  
AC  
XX 08-MAY-2001 (first entry)  
DT  
XX  
DE S. aureus staphylokinase mutagenesis primer 2b.  
XX  
KW Immunogenicity; staphylokinase; variant; stability; mutagenesis;  
KW PCR primer; ss.  
XX  
OS Staphylococcus aureus.  
XX  
PN WO200104287-A1.  
XX  
PD 18-JAN-2001.  
XX  
PE 06-JUL-2000; 2000WO-DK00371.  
XX  
PR 07-JUL-1999; 99DK-0000988.  
PR 27-AUG-1999; 99DK-0001196.  
PR 02-MAR-2000; 2000DK-0000339.  
PR 18-MAY-2000; 2000DK-0000804.  
XX  
PA (MAXY-) MAXYGEN APS.  
XX  
PI Halkier T, Pedersen AH, Okkels JS;  
XX  
DR WPI; 2001-138342/14.  
XX  
PT Producing polypeptides with altered immunogenicity or improved  
PT stability, comprises expressing a diversified nucleotide sequence  
PT population and selecting polypeptides with altered immunogenicity or

PT Improved stability -  
XX  
PS Example 2; Page 62; 83pp; English.  
XX  
CC The invention relates to a method of altering immunogenicity and/or  
CC increasing stability of a polypeptide of interest. The method comprises  
CC (a) expressing a diversified population of nucleotide sequences encoding  
CC a polypeptide of interest; (b) screening the polypeptides expressed for  
CC function, immunogenicity and/or stability; and (c) selecting functional  
CC polypeptides with altered immunogenicity and/or increased stability. The  
CC method is used to improve the properties of polypeptides, in particular  
CC to alter the immunogenicity and/or increase the functional in vivo half-  
CC life of the polypeptide. The method uses a high throughput system that  
CC makes it possible to search several orders of magnitude more polypeptides  
CC than is possible by previously known approaches. This enhances the chance  
CC of finding the optimal variant from the many thousands of variants that  
CC may be produced. Sequences AAC85113-121 represent PCR primers used in a  
CC focused mutagenesis reaction towards introduction of lysine residues in  
CC S. aureus staphylokinase.  
XX  
SQ Sequence 66 BP; 16 A; 11 C; 8 G; 31 T; 0 other;  
alignment\_scores:  
Quality: 84.00 Length: 18  
Ratio: 4.941 Gaps: 0  
Percent Similarity: 94.444 Percent Identity: 94.444  
alignment\_block:  
US-09-728-670-10 x AAC85118/rev ..  
Align seg 1/1 to reverse of: AAC85118 from: 1 to: 66  
119 HisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleGluLys 135  
||||| ||||| ::::::: |||  
65 CATATTAAAAACCCCTGGAAGAACTTAATTACAAAGGTTGTTATAGAAAA 16  
135 sLys 136  
|||  
15 GAAA 12  
seq\_name: /SIDSl/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL19988  
seq\_documentation\_block:  
ID ABL19988 standard; DNA; 24971 BP.  
XX  
AC ABL19988;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11437.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more







```
XX
DR WPI: 2000-587529/55.
XX
PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited mitochondria which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
PS Claim 102; Page 453-484; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited mitochondria which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 134499 BP; 41565 A; 25130 C; 25225 G; 42577 T; 2 other;

alignment_scores:
Quality: 81.00 Length: 138
Ratio: 1.141 Gaps: 8
Percent Similarity: 51.449 Percent Identity: 28.986

alignment_block:
US-09-728-670-10 x AAF22286/rev ..
Align seg 1/1 to reverse of: AAF22286 from: 1 to: 134499

2 SerSerPheAspLys.....GlyLysTyrLysLysGlyAspAspAl 15
|||||:|||||:|||||:|||||:|||||:
128464 TCATCAATAGACGAGCGGGGAGGTGGAATCTCCAAATAGGAGATTAAG 128415
15 aSerTyrPheGluProThrGlyProTyrLeuMetValAsnValThrGlyV 32
:::|||||:|||||:|||||:|||||:|||||:
128414 CAATTTTGAAGAAACTGTGGGAGTAC.....AAGGAAAGAT 128377
32 aLAspSerLysGlyAsnGluLeuLeuSerProHisTyrValGluPhePr 48
:::|||||:|||||:|||||:|||||:|||||:
128376 TGGTCTGCTAAGCTCAATGACACATATGGCTTACAGACAGCCTTCAA 128327
48 oileysProGlyThr..... 54
|||||:|||||:|||||:|||||:|||||:
128326 GACCCCTATTGGCAGACGCTTTCACCTCCTATAGGAAATCCTGTC 128277
55 .LeuThrLysGluLysIleGluTyrTyrValGluTyrPalaleu..... 68
|||||:|||||:|||||:|||||:|||||:
128276 ATTGGCTGTGAA..CTCGATATAAAGCCATGTGGGAGTTAAGCTC 128230
69 .....AspAlaThrAlaTyrLysGluPheArgValAlaGluLeuAs 82
|||||:|||||:|||||:|||||:|||||:
128229 CTGAACTTCGACATTAACACCGCCGAGGAGGAGGCGGTAAATCCAACCTGAA 128180
82 pProSerAlaLysIleGluValThrTyrTyrAspLysAlaLys..LysL 98
|||||:|||||:|||||:|||||:|||||:
128179 CGATCTCAACAGATTCGCTTAGAAGCCTATGAGAGTCCAAATCTACA 128130
98 ysgLugluThrLysSerPhe.....ProilethrGluLysGlyPhe 111
|||||:|||||:|||||:|||||:|||||:
128129 AGGAGCAACCAAGCTTTCATGATAGAAGATAGTCTCAAGAGATTTT 128080
112 ValValProAsp 115
|||||:|||||:|||||:|||||:|||||:
128079 AAGTTGGTGAT 128068

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABLI2301
seq_documentation_block:
ID ABLI2301 standard; cDNA; 9654 BP.
XX
AC ABLI2301;
XX
DT 26-MAR-2002 (first entry)
```

```
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31385.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
DR P-PSDB; ABB68198.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions
XX
PS Claim 1; SEQ ID NO 31385; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 9654 BP; 2677 A; 2311 C; 2518 G; 2148 T; 0 other;

alignment_scores:
Quality: 80.50 Length: 163
Ratio: 1.032 Gaps: 9
Percent Similarity: 47.853 Percent Identity: 27.607

alignment_block:
US-09-728-670-10 x ABLI2301 ..
Align seg 1/1 to: ABLI2301 from: 1 to: 9654

13 AspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetValAsnVa 29
:::|||||:|||||:|||||:|||||:|||||:
2194 GAGGATTCGGGAGGATTTTCA..ACCTCCGTTGATCTCACCATTGCGCT 2240
29 lThrGlyValAspSerLysGlyAsnGluLeuLeuSerPro..... 42
|||||:|||||:|||||:|||||:|||||:
2241 TACGATGTGAATGATAACGCCGCCGAGTTCGAGCTGCCGACTACAGG 2290
43 .HisTyrValGluPheProileLysProGlyThrThrLeuThrLysGlu 58
|||||:|||||:|||||:|||||:|||||:
2291 CGCACAAATGTGAGCAGAGACATTCGCTGGGAGAACAGCATACTTCGCGTT 2340
59 Lys.....IleGluTyrTyrVa 64
|||||:|||||:|||||:|||||:|||||:
2341 AAAGCCATGACTCCGATTCGCGATCAATCGGAAATCGAATATTAGT 2390
64 l.....GluTyrPalaleuAspAla..... 70
|||||:|||||:|||||:|||||:|||||:
```

2391 ATCCGATGATCATTTCCCGCTGGACTCGAACGGAATCATCGTGACACACA 2440  
71 .....ThrAlaTyrLysGluPheArgValVal 79  
:::||||| ||||| ||:::  
2441 AGCAACTGGATGCGGACACACACAMTGCCTATTATGAGTTTCATAGTAGACT 2490  
80 GluLeuAsp.....ProSerAlaLysIleGluValThrTyr 91  
||| ||:::  
2491 GCCAAGGACAAAGGTGAACCGCCCAATCGGAGTGGCTACAGTTGAGT 2540  
91 rTyrAspLysAsnLysLysGluGluThrLysSerPheProIleThrG 108  
||| |||||::: ||||| |||  
2541 TTACACACAAAAACAAGACGATGAGAG.....CCCAAGTTCT 2578  
108 IuLysGlyPheValValProAspLeuSerGluHisIleLysAsnProGly 124  
::: ||::: ||||| ||| |||  
2579 CCCAGCAGTGTATACCCCAATGTGATGAG.....AATGCCGGG 2619  
125 PheAsn..LeuIleThrLysValValIleGluLysLys 136  
||| ||::: ||||| ::| |||  
2620 CCAAAACACTTGTGTTACCACTGTGTTAGCCTCCGACACAAG 2658  
seq\_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:AAQ14806  
seq\_documentation\_block:  
ID AAQ14806 standard; DNA; 3303 BP.  
AC AAQ14806;  
DT 10-FEB-1992 (first entry)  
DE B.thuringiensis toxin/gp64 viral glycoprotein fusion gene pFAV10.  
XX  
KW chimeric; fusion protein; insecticide; AcNPV; Lepidoptera larvae;  
KM midgut targeting; bacterial endotoxin; ss.  
XX  
OS Bacillus thuringiensis var. tenebriosis.  
OS Autographa californica Nuclear Polyhedrosis Virus.  
XX  
PN WO9117254-A.  
XX  
PD 14-NOV-1991.  
XX  
PF 02-MAY-1991; 91WO-US03008.  
XX  
PR 03-MAY-1990; 90US-0518575.  
XX  
PA (REGC ) UNIV OF CALIFORNIA.  
XX  
PI Silvasubramanian N, Federici A;  
XX  
DR WPI; 1991-353775/48.  
DR P-PSDB; AAR15783.  
XX  
PT Extending host range or toxicity of insecticidal proteins - using  
PT protein capable of binding to gut epithelium of insects  
XX  
PS Claim 24; Fig 16; 61pp; English.  
XX  
CC A polylinker was inserted into the XmnI restriction site at the  
CC carboxyl terminus coding region of B.thuringiensis var. tenebriosis  
CC (Bt) toxin. DNA encoding the gp64 viral membrane protein of AcNPV  
CC was operably linked to the Bt toxin coding sequence via the  
CC polylinker. The gp64 gene sequences act as midgut targeting  
CC signals for bacterial endotoxins. Of three different Bt/gp64 gene  
CC fusions that were constructed, pFAV10 was the longest. It was  
CC transformed into E.coli BL21 and expressed. When tested against  
CC Heliothis virescens larvae, pFAV10 showed higher toxicity than  
CC E.coli PSX12T expressing non-fusion Bt protein.  
CC See also AAQ14807 and AAQ14808.  
XX  
SQ Sequence 3303 BP; 1077 A; 672 C; 724 G; 830 T; 0 other;

alignment\_scores:  
Quality: 80.00 Length: 147  
Ratio: 1.250 Gaps: 6  
Percent Similarity: 43.537 Percent Identity: 24.490  
alignment\_block:  
US-09-728-670-10 x AAQ14806 ..  
Align seg 1/1 to: AAQ14806 from: 1 to: 3303  
4 PheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTyr..... 17  
||||||| ::|||  
1663 TTCGATTA...ACGATTAATAAGGAGACACATTAACTATTAATTCAATT 1709  
18 .....PheGluProThrGlyProTyrLeuM 26  
||||| ::||| |||  
1710 TAAATTAGCAAGTTTCAGCACACCATTCGAATTATCAGGGAATAACTTTAC 1759  
26 etValAsnValThrGlyValAspSerLysGlyAsnGluLeuSerPro 42  
::: |||||::: ||| ||::: |||  
1760 AATAAGCGCTCACAGATTA...AGTGTGAGATAAAGTTATATAGAC 1806  
43 HisTyrValGluPheProIleLysPro..... 51  
::||::: ||| |||  
1807 AAA...ATTGAATTGGGATCGATCCCTCTAGAGTCGACCTGCAGCCCAAG 1853  
52 .....GlyThrThrLeuThrLysGluLysI 60  
||| |||||  
1854 CTGTCCGTACAAGATTAAAACTTGACATTTACCCGCCCAAGAAACGC 1903  
60 IeGluTyrTyrValGluTyrPalaleuAspAlaThrAlaTyrLysGluPhe 76  
::: ||||| :::: ||| ||::: |||  
1904 TGCAAAAGGACGTGGAATATCACCTCTGTGAGACGAGACTACACGAAAC 1953  
77 ArgValVal..... 79  
::: |||  
1954 GTGATTATCGGCTACAAGGGTACTACAGGCGTATGCGTACAACGGCGG 2003  
80 .GluLeuAspProSerAlaLysIleGluValThrTyrTyrAspLysAsnL 96  
::: |||||::: ||||| ||| |||  
2004 CTCGCTGATCCCAACACACACGCGTCGAAGAAACCAATGAACCGCTGAATG 2053  
96 yslLysGluGluThrLysSerPheProIleThrGluLys 109  
|||||::: ||| |||  
2054 TGGGCAAGAGGATTTGCTCATGTGAGCATCAGCGACGACAG 2094  
seq\_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAQ99223  
seq\_documentation\_block:  
ID AAQ99223 standard; cDNA to mRNA; 3394 BP.  
AC AAQ99223;  
XX  
DT 06-MAR-1996 (first entry)  
XX  
DE Japanese oyster transglutaminase cDNA.  
XX  
KW Japanese oyster; transglutaminase; gelling agent;  
KW yoghurt; jelly; cheese; fish-paste; calcium ion activation; ds.  
XX  
OS Crassostrea gigas.  
XX  
FH Key Location/Qualifiers  
FH CDS 305..2620  
FT /\*tag= a  
XX  
PN WO9520662-A1.  
XX  
PD 03-AUG-1995.  
XX  
PF 30-JAN-1995; 95WO-JP00117.  
XX





112 ValValProasp 115  
|||  
47874 AAGTTGGTGAT 47885

seq\_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAF22289

seq\_documentation\_block:

ID AAF22289 standard; DNA; 90336 BP.

XX AAF22289;

XX 20-MAR-2001 (first entry)

XX BAC containing repeats from centromeres 1-4 #12.

XX Centromere; michromosome; vector; ds.

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Claim 102; Page 529-549; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

XX (Arabidopsis thaliana) centromere. The constructs are useful for

XX producing stably inherited michrosomes which can serve as vectors for

XX the construction of transgenic plant and animal cells expressing

XX selected proteins such as hormones, enzymes, interleukins, clotting

XX factors, cytokines, antibodies, and growth factors.

XX Sequence 90336 BP; 26524 A; 18837 C; 18625 G; 26350 T; 0 other;

XX Alignment\_scores: Quality: 80.00 Length: 138

XX Ratio: 1.143 Gaps: 8

XX Percent Similarity: 50.725 Percent Identity: 28.986

XX alignment\_block: US-09-728-670-10 x AAF22289 ..

XX Align seg 1/1 to: AAF22289 from: 1 to: 90336

XX 2 SerSerPheAspLys.....GlyLysTyrLysLysGlyAspAspAl 15

XX 60244 TCATCCACAGACGAGCGGCGAGGTGGAATCTCCACACAGGAGATAAAG 60293

XX 15 aSerTyrPheGluProThrGlyProTyrLeuMetValAsnVal.ThrGly 31

XX 60294 CAATTTTAGAGAAACAATGGCAATTACAAGGAAAGATTGGTCTACTAAG 60343

XX 32 ValAspSerLysGlyAsnGluLeuLeuSerProHisTyrValGluPhePr 48

60344 CTCGAT.....GACGCACCTATGGGCTTACAGAAACAGCCTTCAA 60381

48 oileLysProGlyThr..... 54

60382 GACCCCTATTGGAACGACTCTTCAACCTCCTATGGAATACTCTGC 60431

55 .LeuThrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAspAla 70

60432 ATTTGCCCTGTGAA..CTCGAGTATAAGCCATGTGGCGAGTTAAACTC 60478

71 ThrAlaTyr.....LysGluPheArgValValGluLeuAs 82

60479 CTGAACCTTGGCATTAACCCCAAGAGAGAGCCGATTCACACTGAA 60528

82 ProSerAlaLysIleGluValThrTyrTyrAspLysAsnLys..LysL 98

60529 CGATCTCAACGAGATTGCTTAGAAGCTTATGAGAGTCCAAATCTACA 60578

98 ysgIuGluThrLysSerPhe.....ProIleThrGluLysGlyPhe 111

60579 AGGAGCGAACCAAGCTCTTCCATGACACAGAGATGCTCAAGAGATTTT 60628

112 ValValProasp 115

60629 AAGTTGGTGAT 60640

seq\_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA59395

seq\_documentation\_block: ID AAA59395 standard; DNA; 891 BP.

XX AAA59395;

XX 14-NOV-2000 (first entry)

XX DNA encoding a galactanase of Bacillus subtilis.

XX Galactanase; animal feed; galactan; glycosyl hydrolase; textile;

XX detergent; wine; juice; cellulose processing; ss.

XX Bacillus subtilis.

XX Key Location/Qualifiers

XX CDS 1..891

XX /tag= a

XX /product= "galactanase"

XX /note= "no termination codon given"

XX WO200047711-A2.

XX 17-AUG-2000.

XX 08-FEB-2000; 2000WO-DK00052.

XX 11-FEB-1999; 99DK-0000184.

XX 07-JUN-1999; 99DK-0000799.

XX (NOVO ) NOVO NORDISK AS.

XX Bjoernvad ME, Clausen IG, Schuelein M, Bech L, Oestergaard PR;

XX Sjoeholm C;

XX WPI; 2000-565292/52.

XX P-PSDB; AAB07811.

XX Novel methods for modifying animal feed using galactanase and novel

XX galactanase enzymes useful for modifying animal feed -

XX Claim 22; Page 65-66; 77pp; English.

XX The present sequence encodes a Bacillus galactanase enzymes. The

XX enzyme comprises at least one of the consensus sequences AAB07802-07,

CC and is used in the method of the invention. The specification  
CC describes a method for modifying animal feed using galactanase.  
CC Galactanase enzymes degrade galactans. The galactanases of  
CC the invention are glycosyl hydrolases. The galactanase enzymes are  
CC useful for the modification of animal feed and in the textile,  
CC detergent, wine and juice and cellulose processing industries.

Sequence 891 BP; 261 A; 183 C; 263 G; 184 T; 0 other;

alignment\_scores:                      Length:    116  
                    Quality:    79.50  
                    Ratio:    1.242  
Percent Similarity: 55.172    Percent Identity: 29.310

alignment\_block:

US-09-728-670-10 x AAC59395 ..

Align seg 1/1 to: AAC59395 from: 1 to: 891

19 GluProThrGlyProTyrLeu...MetValAsnValThrGlyValAspSe 34  
   ||| |||||    ::    |||    ::    ::    ::  
400 GAAACACACAGGAGATTGCCACAGATGTGCCAATATTATGAAGGAAG 449  
   |||    ::    ::    ::    ::    ::    ::  
34 rlysglyAsnGluLeuSerProHisTyrValGluPheProIleLysP 51  
   |:::    ::|||:::    |||:::    ::    ::    ::  
450 CCGACGGTCAGGTGGTGGCCCTGCAATTTACCAAT...CTGAAACGG 496  
   ||| |||||::: |||::: |||::: |||    ::|||  
51 roglyThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAla 67  
   ||| |||||::: |||::: |||::: |||    ::|||  
497 CTGGAAGACACTCAGCAAAACAAAGTGATATGATGTG...TTTGT 543  
68 LeuAspAlaThr.....AlaTyrLysGluPheArgVa 78  
   ::|||    ::    ::    ::    ::    ::  
544 AGCCATGGCACATTACAAATTTGACCTCCGTGCTGAAGCTGTAAAGT 593  
78 lValGluLeuAspProSerAlaLysIleGluValThrTyrTrpAspLysA 95  
   |:::    |||:::    ::    |||||    ::|||  
594 CATG.....GTGGCGAGACATCGTACACCTATACCGCTGAGA 631  
95 sn.....LysLysLysGluGluThrLysSerPheProIleThrGluLys 109  
   ||    |||:::    ::|||    ::|||    ::|||  
632 ATACAGCACCAAAAGCGGGCAGACGTTGCCATATCCAACTACT.... 675  
110 GlyPheValValProAspLeuSerGluHisIleLysAsnProGlyPhe 125  
   ::|||    |||:::    |||    ::    |||    |||||  
676 .....GCAGTAAGAGATGTAATGAGGACAGTGGCGAATACGGGCTTC 717

seq\_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAC85116

seq\_documentation\_block:

ID AAC85116 standard; DNA; 42 BP.

AC AAC85116;

DT 08-MAY-2001 (first entry)

DE S. aureus staphylokinase mutagenesis primer 1d (wt).

KW Immunogenicity; staphylokinase; variant; stability; mutagenesis;

KW PCR primer; ss.

OS Staphylococcus aureus.

PN WO200104287-A1.

PD 18-JAN-2001.

PF 06-JUL-2000; 2000WO-DK00371.

PR 07-JUL-1999; 99DK-0000988.

PR 27-AUG-1999; 99DK-0001196.

PR 02-MAR-2000; 2000DK-0000339.

PR 18-MAY-2000; 2000DK-0000804.

XX (MAXY-) MAXYGEN APS.

XX Halkier T, Pedersen AH, Okkels JS;

XX WPI; 2001-138342/14.

XX Producing polypeptides with altered immunogenicity or improved

PT stability, comprises expressing a diversified nucleotide sequence

PT population and selecting polypeptides with altered immunogenicity or

PS Example 2; Page 62; 83pp; English.

XX The invention relates to a method of altering immunogenicity and/or  
CC increasing stability of a polypeptide of interest. The method comprises  
CC (a) expressing a diversified population of nucleotide sequences encoding  
CC a polypeptide of interest; (b) screening the polypeptides expressed for  
CC function, immunogenicity and/or stability; and (c) selecting functional  
CC polypeptides with altered immunogenicity and/or increased stability. The  
CC method is used to improve the properties of polypeptides, in particular  
CC to alter the immunogenicity and/or increase the functional in vivo half-  
CC life of the polypeptide. The method uses a high throughput system that  
CC makes it possible to search several orders of magnitude more polypeptides  
CC than is possible by previously known approaches. This enhances the chance  
CC of finding the optimal variant from the many thousands of variants that  
CC may be produced. Sequences AAC85113-121 represent PCR primers used in a  
CC focused mutagenesis reaction towards introduction of lysine residues in  
S. aureus staphylokinase.

Sequence 42 BP; 14 A; 9 C; 12 G; 7 T; 0 other;

alignment\_scores:                      Length:    14  
                    Quality:    79.00  
                    Ratio:    5.643  
Percent Similarity: 100.000    Percent Identity: 100.000

alignment\_block:

US-09-728-670-10 x AAC85116 ..

Align seg 1/1 to: AAC85116 from: 1 to: 42

10 LysLysGlyAspAspAlaSerTyrPheGluProThrGlyPro 23  
   |||||    |||||    |||||    |||||    |||||  
1 AAAAAAGCGCATGACGCGAGTTATTGAAACCAACAGCCCG 42

seq\_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:AAN60718

seq\_documentation\_block:

ID AAN60718 standard; DNA; 437 BP.

AC AAN60718;

DT 23-OCT-1991 (first entry)

DE Fragment of sak gene in novel compound plasmid.

KW Colibacillus; sak promoter; Psak; SIGsak; ss.

XX Staphylococcus aureus.

OS

FH Key Location/Qualifiers

FT promoter 191..220

FT RBS 299..303

FT CDS 311..634

FT sig\_peptide 311..391

FT sig\_peptide d





```

seq_documentation_block:
ID   AAS46743 standard; DNA; 14147 BP.
XX
AC   AAS46743;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #467.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer
XX
PS   Claim 1; SEQ ID No 467; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 14147 BP; 3621 A; 225 C; 3316 G; 7185 T; 0 other;

```

alignment_scores:		
Quality:	78.00	Length: 119
Ratio:	1.238	Gaps: 6
Percent Similarity:	52.941	Percent Identity: 26.050

alignment\_block:  
US-09-728-670-10 x AAS46743/rev ..

Align seg 1/1 to reverse of: AAS46743 from: 1 to: 14147

28 AsnValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHisTyr 44

13492 AATCTCAAAAATTAATCTTACCAACAC...TTAAATTAACAAA 13449

44 rvaIglupheproIlelysproglythrThrleuthrlysglulysilec 61

13448 AATAATTTCCTAAACCT.....CCAAAAAACGC 13414

61 LUTYRTYRValGluTRPalaLeuaspalaThrAlaTYRLys..... 74

13413 AACCTACCAATACCTTAATTTCACCTAAACCAATATAAACTTCTAAC 13364

75 .....GlupheargValValGlu...LeuAspProSerAlaLysIleG1 88

13363 CTACAAACTATATAATTAACAATATATTATTTTAACACACTAAATTAT 13314

88 uvalthrtYrTyrAspLysAsnLysLysGluGluThrLysSerPhe. 104

13313 AATTATTATTACACAAATAAAACTAATTCATCCTCCTCATTTA 13264

105 .....ProleThrgluysgly 110

13263 AATATTATCTATCTTCTACCATATCAATATCAAAATTTTCAAAAAATTT 13214

111 PheValValPro.....AspLeuSerGluHisIleLysAsnPr 123

13213 TTTTATTACCAACCTTACAAAACTACTCACTCCACGTAAACTTCC 13164

123 oglyphe 125

13163 TTCTTTC 13157

seq\_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV72417

seq\_documentation\_block:

ID AAV72417 standard; DNA; 10220 BP.

AC AAV72417;

DT 03-AUG-1999 (first entry)

DE P. Luminescens pTetLuxI LuxCDABE DNA.

KW Firefly; luciferase; tetracycline; transcriptional control; TetR; TetA;

KW insect; medicine; dosage; cheese production; antibiotic; foodstuff;

XX

**XXII**

FT	CDS	3634..5085
----	-----	------------

FT /product= "Lux C"

/\*tag= b

FT	CDS	6069..7151
----	-----	------------

/product= "Lux A"

/\*tag= d

FT	8256..9440
CDS	

/product= "Lux E"

PN W099925866-A1.

```
XX 27-MAY-1999.
PD
XX 11-NOV-1998; 98WO-FI00873.
PF
XX 14-NOV-1997; 97FI-0004235.
PR
XX
XX (KARP/) KARP M.
PA (KORP/) Korpela M.
PA (KURI/) Kurittu J.
XX
PI Karp M, Korpela M, Kurittu J;
XX
DR WPI; 1999-338015/28.
DR P-PSDB; AAY08524, AAY08525, AAY08526, AAY08527, AAY08528;.
XX
XX Assaying for tetracycline using recombinant prokaryotic cells
XX
XX Claim 3; Page 28-37; 67pp; English.
XX
XX This invention describes a novel tetracycline assay that uses recombinant
CC prokaryotic cells comprising a luciferase gene under the transcriptional
CC control of a tetracycline repressor and tetracycline promoter and
CC involves the detection of luminescence emitted from the cells. The assay
CC can be used to distinguish tetracycline form other microbial agents. The
CC invention also describes a novel plasmid comprising either the luxCDABE
CC genes, a tetracycline repressor (Tetr) and a tetracycline promoter (Tetr)
CC from Tn10, or the insect luciferase gene, a tetracycline repressor (tetr)
CC and a tetracycline promoter (Tetr) from Tn10. The tetracycline assay
CC method can be used for the determination of tetracycline in a sample,
CC e.g. to study the dosage and penetration of the medicine. The method can
CC also be used to test cheese production, as cheese making bacteria are not
CC able to work in the presence or concentration of tetracycline. The method can
CC to determine the presence or concentration of antibiotics in foodstuffs,
CC e.g. for allergic people. The present assay method does not rely on the
CC growth of microbes as do conventional tests, and so is much more rapid.
CC The present assay is also more sensitive, as even a small amount of
CC luminescence can be detected.
XX
XX Sequence 10220 BP; 3127 A; 1832 C; 2255 G; 3006 T; 0 other;
```

```
alignment_scores:
Quality: 75.50 Length: 130
Ratio: 1.280 Gaps: 7
Percent Similarity: 45.385 Percent Identity: 25.385
```

alignment\_block:

US-09-728-670-10 x AAV72417 ..

Align seg 1/1 to: AAV72417 from: 1 to: 10220

```
18 PheGluProHrGlyProTyrLeuMetValAsnValThrGlyValAspSe 34
::: ||| ||||| ||::: |||
7652 TATACGCCAGGCGGACCTCGGAATATGTACAGCAACC..... 7690
34 rLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLysP 51
||| ||| |||||::: |||
7691 .....AGTCATCATATGTGTGAGTGGCGGCCCAAA 7721
51 roGlyThrThrLeu.....ThrLysGluLysIle 60
||| ||| ::: |||
7722 AAGGTATTCCTCATCTTTAAGTGGATGATCTAATGATGTTAGATAT 7771
61 GluTyrTyrValGluTyrAlaLeuAspAlaThrAlaTyrLysGluPheAr 77
||||| :::: ||| ||| :::
7772 GAATATGCTGAAGATATAAAGCCGTTGCGGATAATAT...GACGTTGA 7818
77 gValValGluLeuAspProSerAlaLysIleGluValThrTyr...TyrA 93
::: ||::: ||| ::: ||| ||::: |||
7819 CCTATCAGAGATAGACCATCATGTTAATGATATTAAGTTAACAACGAC 7868
93 sPLySAsnLysLysGluGluThrLysSerPheProIleThrGluLys 109
```

```
||||| ||::: |||||::: |||
7869 ATAGTAATAAGCTAAACAAGAGACGCGTGCAATT..... 7903
110 GlyPheValProAspLeuSerGluHisIle.....LysAs 122
::: ||::: |||
7904 .....ATTAGTATATGTTCTTGAATGCACCCCTAA 7935
122 nProGlyPhe...AsnLeuIleThrLysValValIleGlu 134
||| ::: ||| ||| ::: |||
7936 TGAATAATTTCGAATAAATACTGACAGAAATATTCACAGAA 7975
```

seq\_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAC85113

seq\_documentation\_block:

ID AAC85113 standard; DNA; 42 BP.

AC AAC85113;

DT 08-MAY-2001 (first entry)

DE S. aureus staphylokinase mutagenesis primer 1a.

KW Immunogenicity; staphylokinase; variant; stability; mutagenesis;

KW PCR primer; ss.

OS Staphylococcus aureus.

PN WO200104287-A1.

PD 18-JAN-2001.

PF 06-JUL-2000; 2000WO-DK00371.

XX 07-JUL-1999; 99DK-0000988.

PR 27-AUG-1999; 99DK-0001196.

PR 02-MAR-2000; 2000DK-0000339.

PR 18-MAY-2000; 2000DK-0000804.

XX (MAXY-) MAXYGEN APS.

PI Halkier T, Pedersen AH, Okkels JS;

DR WPI; 2001-138342/14.

PT Producing polypeptides with altered immunogenicity or improved

PT stability, comprises expressing a diversified nucleotide sequence

PT population and selecting polypeptides with altered immunogenicity or

PS improved stability -

XX Example 2; Page 61; 83pp; English.

XX The invention relates to a method of altering immunogenicity and/or

XX increasing stability of a polypeptide of interest. The method comprises

XX (a) expressing a diversified population of nucleotide sequences encoding

XX a polypeptide of interest; (b) screening the polypeptides expressed for

XX function, immunogenicity and/or stability; and (c) selecting functional

XX polypeptides with altered immunogenicity and/or increased stability. The

XX method is used to improve the properties of polypeptides, in particular

XX to alter the immunogenicity and/or increase the functional in vivo half-

XX life of the polypeptide. The method uses a high throughput system that

XX makes it possible to search several orders of magnitude more polypeptides

XX than is possible by previously known approaches. This enhances the chance

XX of finding the optimal variant from the many thousands of variants that

XX may be produced. Sequences AAC85113-121 represent PCR primers used in a

XX focused mutagenesis reaction towards introduction of lysine residues in

XX S. aureus staphylokinase.

XX Sequence 42 BP; 15 A; 9 C; 12 G; 6 T; 0 other;

alignment\_scores:

Quality: 75.00 Length: 14  
Ratio: 5.357 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 92.857

alignment\_block:

US-09-728-670-10 x AAC85113 ..

Align seg 1/1 to: AAC85113 from: 1 to: 42

10 LysLysGLYAspAspAlaSerTyrPheGluProThrGlyPro 23  
|||||  
1 AAAAAGGCGATGACGCGAGTATTGACCAACAGGCCG 42

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:ABL34443

seq\_documentation\_block:

ID ABL34443 standard; DNA; 9899 BP.

XX ABL34443;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2416.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.

XX Homo sapiens.

PN WO200200928-A2.

XX 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -

PS Claim 1; SEQ ID NO 2416; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.

XX Sequence 9899 BP; 2369 A; 359 C; 2471 G; 4700 T; 0 other;

alignment\_scores:

Quality: 74.50 Length: 138  
Ratio: 1.146 Gaps: 6  
Percent Similarity: 47.101 Percent Identity: 26.087

alignment\_block:

US-09-728-670-10 x ABL34443/rev ..

Align seg 1/1 to reverse of: ABL34443 from: 1 to: 9899

37 AsnGluLeuLeuSerProHisTyrValGluPhe..... 47

2821 AACACCCGACACCCGACACTACCTTCACAAACACGACACT 2772

48 .....ProIleLysProGlyThrThrL 55

2771 CATTAATAATACGACCAAAACCAATAATTAATTAATCCCAAAACACTA 2722

55 eu.....ThrLysGluLysIleGluTyrTyrValGluTrrPalaleuasp 69

2721 TTTTAAATCTAAACGCTTAATCGACAAATAATTAATTAATCTATTATA 2672

70 AlaThrAlaTyrLys.....GluPheArgValValGluLeuaspProse 84

2671 CAACCAATAATTAATACTTTAACTTTCAATAATTAATTAATAATAAAACAA 2622

84 rAlaLysIleGluValThrTyrTyrAspLysAsnLysLys..... 97

2621 ACCCAAAATCCTTATATCTACAAAAAAATAAACTTCTACCCCTA 2572

98 .....LysGluGluThrLysSer 103

2571 AAAAATCTCTTCGATAATATTAATTAATTAATAAACTACTAAAC 2522

104 Phe.....ProIleThrGluLysGlyPheValValPro..... 114

2521 TTTTAAACCACTTATCACAATATTAATTAATTAATTAATTAATTAATTA 2472

115 AspleuSerGluHisIleLysAsnProGlyPheAsnLeuIleThrLysV 131

2471 CAATACTCTCAACGTCAAACTACTAAATAATCAATAATTAATTAATTA 2422

131 alValIleGluLys 135

2421 TTAATATACTTAA 2408

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAC47096

seq\_documentation\_block:

ID AAC47096 standard; DNA; 1645 BP.

XX AAC47096;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 52550.

XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.



```
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

## alignment\_scores:

Quality:	74.00	Length:	133
Ratio:	1.088	Gaps:	7
Percent Similarity:	51.128	Percent Identity:	28.571

## alignment\_block:

US-09-728-670-10 x AAC47096 ..

Align seg 1/1 to: AAC47096 from: 1 to: 1645

```
6 LysGlyLysTyrLysGlyAspAlaSerTyrPheGluProThrG1 22
   |||::: ||| ||| ||| |||::: |||
768 AAGCCCTTCCAGAACCAGCGGCTTGATTAACCAACCGGTTATCCGGTTGG 817
22 yProTyrLeuMetValAsnValThrGlyValAspSerLysGlyAsnGlu. 38
   ||| |||::: ||| |||::: ||| |||::: |||
818 ACCG.....TTGGTTAACAATGGTAAGCAAGAGGCTAAGCAACCGAAG 861
39 .....LeuLeuSerProHis 43
   ||| |||
862 AGTCGAAATGTTTAAAGTGGTGATTAACCAACCGCGCTCGGTTGTTTA 911
44 TyrValGluPheProIleLysProGlyThrThrLeuThrLysGluLysI1 60
   ||| |||::: ||| ||| ||| ||| ||| |||::: |||
912 TATGTGTCTCTT.....GGTAGTGGCGGTACCCCTCACATGTGAGCAGCT 955
60 eGluTyrTyrValGluTrpAlaLeuAsp...AlaThrAlaTyrLysGluP 76
   ::| ||| ||| ||| ||| ||| |||::: |||
956 CAAT.....GAGCTTGCTCTTGGTCTTGACAGATAGTACCAACGCT 996
76 heArgValValGluLeuAspProSerAlaLysIleGluValThrTyrTyr 92
   || ||| |||::: ||| |||::: ||| |||::: |||
997 TTCTTTGGGTCATACGAAGTCTTAGTGGATCGCTAATTCGTCTATTATT 1046
93 AspLysAsnLysLysLysGluGluThrLysSerPheProIleThrGluLy 109
   |||::: |||::: |||::: |||::: |||::: |||
1047 GATTCACATAGCCCAACAGAT.....CCATGACA..... 1076
109 sGlyPheValValProAspLeuSerGluHisIleLysAsnProGlyPhe 125
   |||::: |||::: |||::: |||::: |||::: |||
1077 ....TTTATTACCACCGGATTTTATAGACCGGACTAAAAAAGAGGTTT 1121
```



OM of: US-09-728-670-10 to: Issued\_Patents\_NA:\* out\_format : pfs

Date: Sep 1, 2002 4:30 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

#### Command line parameters:

```
-MODEL=frimet-p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09728670/runat_30082002_084913_5318/app_query.fasta.1.196
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=50 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09728670_@CGN1_1_45 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

#### Search information block:

```
Query: US-09-728-670-10
Query length: 136
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 46.280000
```

#### score\_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-256-261-5 +				715.00	1710.80	1.3e-87
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-852-299-5 +				715.00	1710.80	1.3e-87
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-852-299-5 +				715.00	1710.80	1.3e-87
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-256-261-3 +				711.00	1701.13	4.6e-87
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-852-299-3 +				711.00	1701.13	4.6e-87
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-256-261-1 +				695.00	1662.46	6.5e-85
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-852-299-1 +				695.00	1662.46	6.5e-85
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-256-261-16 +				695.00	1652.50	2.3e-84
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-852-299-16 +				695.00	1652.50	2.3e-84
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-256-261-11 +				689.00	1647.96	4.2e-84
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-852-299-11 +				689.00	1647.96	4.2e-84
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-256-261-13 +				687.00	1643.13	7.8e-84
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-852-299-13 +				687.00	1643.13	7.8e-84
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-256-261-7 +				643.00	1537.61	5.9e-78
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-852-299-7 +				643.00	1537.61	5.9e-78
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-256-261-9 +				620.00	1482.37	7.0e-75
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-852-299-9 +				620.00	1482.37	7.0e-75
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-256-261-9 -				84.50	150.61	1.06
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-852-299-9 -				84.50	150.61	1.06
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-09-321-831-1 -				81.50	154.39	0.6534
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-525-654A-136 +				80.00	152.91	0.7905
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-09-251-645-11 +				73.50	110.61	179.31
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-480-481-3 +				73.00	148.91	1.32
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-480-481-4 +				73.00	148.91	1.32
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-525-654A-2 +				73.00	140.22	4.02
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-525-654A-4 +				73.00	140.21	4.03
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-469-005A-9 +				73.00	125.93	25.14
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-261-907-1 +				73.00	125.85	25.39
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-282-305-9 -				70.50	138.39	5.09
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-532-656-7 +				69.00	140.65	3.81
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-795-475-4 +				69.00	136.82	6.22
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-064-693A-22 +				69.00	136.49	6.49
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-064-693A-26 +				69.00	123.14	35.96
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-960-780-26 +				68.50	135.87	7.03
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-073-898-26 +				68.50	135.87	7.03
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-257-073-10 +				68.50	120.45	50.76
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-198-956-1 +				68.00	133.00	10.15
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-131-648-4 +				67.00	125.82	25.51
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-968-563-7 +				67.00	122.21	40.52
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-969-683A-7 +				67.00	122.21	40.52
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-297-928-3 +				67.00	122.21	40.52
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-371-913A-10 +				66.50	133.29	9.78

```
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-313-930-1 + 66.00 124.33 30.86 2104
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-147-208-26 - 66.00 107.58 264.73 9639
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-978-741-5 - 66.00 105.84 330.71 11284
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-852-481-1 + 65.50 126.39 23.70 1564
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-232-079-1 + 65.50 120.48 50.58 2675
```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-256-261-5

#### seq documentation block:

```
; Sequence 5, Application US/08256261
; Patent No. 5801037
;
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schloft, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; NUMBER OF SEQUENCES: 40
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/256,261
;
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: strain 23, genomic DNA
;
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414
; OTHER INFORMATION: /note= "stop codon"
;
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4..411
; OTHER INFORMATION: /product= "mature protein"
;
; US-08-256-261-5
```

#### alignment\_scores:

```
Quality: 715.00 Length: 136
Ratio: 5.257 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

#### alignment\_block:

```
US-09-728-670-10 x US-08-256-261-5 ..
```

Align seg 1/1 to: US-08-256-261-5 from: 1 to: 414

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17  
|||||  
4 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGAGTTA 53  
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
54 TTTGAACCAACAGAGCCCGTATTGATGTGAATGTGACTGAGTTGATA 103  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
|||||  
104 GTAAAGGAATGAATGCTATCCCTCATTTATGTCGAGTTCCCTATTAA 153  
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPAl 67  
|||||  
154 CCTGGACTACACTTACAAAAGAAAATGAAATGAACTATGTCGAAATGGGC 203  
67 aleuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspProS 84  
|||||  
204 ATTAGATCGACAGCATATAAGAGTTAGAGTAGTTGAATTAGATCCAA 253  
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100  
|||||  
254 GCGCAAGATCGAAGTCACCTATTATGATGAATAAGAAAAAGAGAGAA 303  
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117  
|||||  
304 ACGAAGTCTTCCCTATAACAGAAAAAGGTTTGTGTCCAGATTATC 353  
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134  
|||||  
354 AGAGCATATTAAAAACCTGGATTCACACTTAATTACAAAGGTTGTATAG 403  
134 LuLysLys 136  
|||||  
404 AAAAGAA 411

seq\_name: /cgn2\_6/ptodata/1/lna/6A\_COMB.seq: US-08-852-299-5

seq\_documentation\_block:  
; Sequence 5, Application US/08852299  
; Patent No. 6010897  
; GENERAL INFORMATION:  
; APPLICANT: Behnke, Detlef  
; APPLICANT: Schlotz, Bernhard  
; APPLICANT: Albrecht, Sybille  
; APPLICANT: G hrs, Karl-Helz  
; APPLICANT: Hartmann, Manfred  
; TITLE OF INVENTION: Expression of signal-peptide-free  
; NUMBER OF INVENTIONS: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,299  
; FILING DATE: 17-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/256,261  
; FILING DATE:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 414 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "recombinant DNA"  
ORIGINAL SOURCE:  
ORGANISM: Staphylococcus aureus  
STRAIN: strain 23, genomic DNA  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..3  
OTHER INFORMATION: /note= "start codon"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..414  
FEATURE:  
NAME/KEY: -  
LOCATION: 412..414  
OTHER INFORMATION: /note= "stop codon"  
FEATURE:  
NAME/KEY: mat-peptide  
LOCATION: 4..411  
OTHER INFORMATION: /product= "mature protein"  
US-08-852-299-5

alignment\_scores:  
Quality: 715.00 Length: 136  
Ratio: 5.257 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-728-670-10 x US-08-852-299-5 ..

Align seg 1/1 to: US-08-852-299-5 from: 1 to: 414

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17  
|||||  
4 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGAGTTA 53  
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
54 TTTGAACCAACAGAGCCCGTATTGATGTGAATGTGACTGAGTTGATA 103  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
|||||  
104 GTAAAGGAATGAATGCTATCCCTCATTTATGTCGAGTTCCCTATTAA 153  
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPAl 67  
|||||  
154 CCTGGACTACACTTACAAAAGAAAATGAAATGAACTATGTCGAAATGGGC 203  
67 aleuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspProS 84  
|||||  
204 ATTAGATCGACAGCATATAAGAGTTAGAGTAGTTGAATTAGATCCAA 253  
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100  
|||||  
254 GCGCAAGATCGAAGTCACCTATTATGATAAGAAATGAAAAAGAGAGAA 303  
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117  
|||||  
304 ACGAAGTCTTCCCTATAACAGAAAAAGGTTTGTGTCCAGATTATC 353  
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134  
|||||  
354 AGAGCATATTAAAAACCTGGATTCACACTTAATTACAAAGGTTGTATAG 403  
134 LuLysLys 136  
|||||  
404 AAAAGAA 411



seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-256-261-3

seq\_documentation\_block:

; Sequence 3, Application US/08256261

; Patent No. 5801037

; GENERAL INFORMATION:

; APPLICANT: Behnke, Detlef

; APPLICANT: Schlot, Bernhard

; APPLICANT: Albrecht, Sybille

; APPLICANT: G hrs, Karl-Heinz

; APPLICANT: Hartmann, Manfred

; TITLE OF INVENTION: Expression of signal-peptide-free

; TITLE OF INVENTION: staphylokinases

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Neave

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/256, 261

; FILING DATE:

; CLASSIFICATION: 435

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 414 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "recombinant DNA"

; ORIGINAL SOURCE:

; ORGANISM: Staphylococcus aureus

; STRAIN: Phage phiC

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..414

; FEATURE:

; NAME/KEY: -

; LOCATION: 1..3

; OTHER INFORMATION: /note= "start codon"

; FEATURE:

; NAME/KEY: -

; LOCATION: 412..414

; OTHER INFORMATION: /note= "stop codon"

; NAME/KEY: mat\_peptide

; LOCATION: 4..411

; OTHER INFORMATION: /product= "mature protein"

; US-08-256-261-3

alignment\_scores:

Quality: 711.00 Length: 136

Ratio: 5.228 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.265

alignment\_block:

US-09-728-670-10 x US-08-256-261-3 ..

Align seg 1/1 to: US-08-256-261-3 from: 1 to: 414

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTyr 17

4 TCAAGTTCATTCGACAAAGAAATATATAAAAGGCGATGCGCGAGTTA 53

seq\_name: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-08-852-299-3

seq\_documentation\_block:

; Sequence 3, Application US/08852299

; Patent No. 6010897

; GENERAL INFORMATION:

; APPLICANT: Behnke, Detlef

; APPLICANT: Schlot, Bernhard

; APPLICANT: Albrecht, Sybille

; APPLICANT: G hrs, Karl-Heinz

; APPLICANT: Hartmann, Manfred

; TITLE OF INVENTION: Expression of signal-peptide-free

; TITLE OF INVENTION: staphylokinases

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Neave

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/852, 299

; FILING DATE: 17-MAY-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/256, 261

; FILING DATE:

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 414 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "recombinant DNA"

```

; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Phage Ph1C
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: -
; LOCATION: 412..414
; OTHER INFORMATION: /note= "stop codon"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4..411
; OTHER INFORMATION: /product= "mature protein"
; US-08-852-299-3
```

```
alignment_scores:
    Quality: 711.00      Length: 136
    Ratio: 5.228         Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 99.265
```

```
alignment_block:
US-09-728-670-10 x US-08-852-299-3 ..
```

```
Align seg 1/1 to: US-08-852-299-3 from: 1 to: 414
```

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17
  |||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAAGGCGATGACGCGAGTTA 53
  |||||
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
  |||||
54 TTTTGAACCAACAGGCGCCGATTTGATGCTAAATGTCAGCTGAGTTGATG 103
  |||||
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
  ::|||
104 GTAAAGGAATGAATGCTATCCCTCATTTATGTCGAGTTTCCCTATTAAA 153
  |||||
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPAl 67
  |||||
154 CCTGGACTACACTTACAAAAGAAAATGAATGACTATGTCGAATGGGC 203
  |||||
67 aLeuAspAlaThrAlaTyrLysGluPheArgValAlGluLeuAspProS 84
  |||||
204 ATTAGATGCAGACAGCATATAAAGAGTTTAGAGTAGTGAATTAGATCCAA 253
  |||||
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
  |||||
254 GCGCAAGATCGAAGTCACCTTATATGATAAGATAAGAAAAAGAGAA 303
  |||||
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
  |||||
304 ACGAAGTCTTCCCTATAACAGAAAAAGGTTTGTGTCGCCAGATTATC 353
  |||||
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
  |||||
354 AGAGCATATTAAAAACCTCGATTCAACTTAATTACAAAGGTTGTTATAG 403
  |||||
134 LuLysLys 136
  |||||
404 AAAAGAAA 411
```

```
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-256-261-1
```

```
seq_documentation_block:
; Sequence 1, Application US/08256261
; Patent No. 5801037
; GENERAL INFORMATION:
```

```

; APPLICANT: Behnke, Detlef
; APPLICANT: Schlot, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/256,261
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Phage 42D
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: -
; LOCATION: 412..414
; OTHER INFORMATION: /note= "stop codon"
; NAME/KEY: mat_peptide
; LOCATION: 4..411
; OTHER INFORMATION: /product= "mature protein"
; US-08-256-261-1
```

```
alignment_scores:
    Quality: 695.00      Length: 136
    Ratio: 5.148         Gaps: 0
    Percent Similarity: 99.265    Percent Identity: 97.794
```

```
alignment_block:
US-09-728-670-10 x US-08-256-261-1 ..
```

```
Align seg 1/1 to: US-08-256-261-1 from: 1 to: 414
```

```

1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAlaSerTy 17
  |||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAAGGCGATGACGCGAGTTA 53
  |||||
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
  |||||
54 TTTTGAACCAACAGGCGCCGATTTGATGCTAAATGTCAGCTGAGTTGATG 103
  |||||
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
  ::|||
```

```
104 GTAAAGCAATGATTCCTATCCCTCGTTATGTCGAGTTTCTATTAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
   |||||
154 CCTGGGACTACACTTACAAAGAAAAAATTGAATACTATGTCGAATGGC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspProS 84
   |||||
204 ATTAGATCGACAGCATATAAGATTAGAGTAGTAGTGAATTAGATCCAA 253
84 eraAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
   |||||
254 GCGCAAGATCGAAGTCACCTATATGATAGAATAGAAAAAGAGAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValProAspLeuSe 117
   |||||
304 ACGAAGCTTTCCTATAACAGAAAAAGTTTGTGTGCCAGATTATTC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
   |||||
354 AGAGCATATTAAAAACCCGTGATTCACCTAATTACMAAGTTGTATAG 403
134 LuLysLys 136
   |||||
404 AAAAGAAA 411
```

seq\_name: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-08-852-299-1

seq\_documentation\_block:

```
; Sequence 1, Application US/08852299
; Patent No. 6010897
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlott, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,299
; FILING DATE: 17-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/256,261
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Phage 42D
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414
```

```
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
; OTHER INFORMATION: /note= "start codon"
; FEATURE:
; NAME/KEY: -
; LOCATION: 412..414
; OTHER INFORMATION: /note= "stop codon"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4..411
; OTHER INFORMATION: /product= "mature protein"
; US-08-852-299-1
```

```
alignment_scores:
  Quality: 695.00      Length: 136
  Ratio: 5.148        Gaps: 0
  Percent Similarity: 99.265  Percent Identity: 97.794
```

alignment\_block:  
US-09-728-670-10 x US-08-852-299-1 ..

Align seg 1/1 to: US-08-852-299-1 from: 1 to: 414

```
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTyr 17
   |||||
4 TCAAGTTCATTCGACAAAGAAAAATATATAAAAGCGATGACCGCGACTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
   |||||
54 TTTTGAACCAACAGCCCGCTATTGTGATGTAATGTGACTGGAGTTGATG 103
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
   ::|||
104 GTAAAGAAATGAATGCTATCCCTCGTATGTCGAGTTTCTATTAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
   |||||
154 CCTGGGACTACACTTACAAAGAAAAAATTGAATACTATGTGCAATGGGC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspProS 84
   |||||
204 ATTAGATCGACAGCATATAAGAGTTTAGAGTAGTGAATTAGATCCAA 253
84 eraAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
   |||||
254 GCGCAAGATCGAAGTCACCTATATGATAGAATAGAAAAAGAGAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValProAspLeuSe 117
   |||||
304 ACGAAGCTTTCCTATAACAGAAAAAGTTTGTGTGCCAGATTATTC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
   |||||
354 AGAGCATATTAAAAACCCGTGATTCACCTAATTACMAAGTTGTATAG 403
134 LuLysLys 136
   |||||
404 AAAAGAAA 411
```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-256-261-16

seq\_documentation\_block:

```
; Sequence 16, Application US/08256261
; Patent No. 5801037
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlott, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Heinz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
```

;; TITLE OF INVENTION: staphylokinases  
;; NUMBER OF SEQUENCES: 40  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Fish & Neave  
;; STREET: 1251 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: USA  
;; ZIP: 10020  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/256,261  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; INFORMATION FOR SEQ ID NO: 16:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1023 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; ORIGINAL SOURCE:  
;; ORGANISM: Staphylococcus aureus  
;; STRAIN: phage 42D  
;; IMMEDIATE SOURCE:  
;; CLONE: Plasmid DB17  
;; FEATURE:  
;; NAME/KEY: -35\_signal  
;; LOCATION: 224..229  
;; FEATURE:  
;; NAME/KEY: -10\_signal  
;; LOCATION: 248..253  
;; FEATURE:  
;; NAME/KEY: -  
;; LOCATION: 332..336  
;; OTHER INFORMATION: /note= "Shine-Dalgarno sequence"  
;; FEATURE:  
;; NAME/KEY: sig\_peptide  
;; LOCATION: 344..415  
;; FEATURE:  
;; NAME/KEY: mat\_peptide  
;; LOCATION: 416..832  
;; OTHER INFORMATION: /product= "mature protein SAK42D"  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 344..835  
;; US-08-256-261-16  
  
alignment\_scores:  
Quality: 695.00 Length: 136  
Ratio: 5.148 Gaps: 0  
Percent Similarity: 99.265 Percent Identity: 97.794  
  
alignment\_block:  
US-09-728-670-10 x US-08-256-261-16 ..  
Align seg 1/1 to: US-08-256-261-16 from: 1 to: 1023  
  
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17  
|||||  
425 TCAAGTTCATTCGACAAAGAAATATATAAAAGCGATGACGCGAGTTA 474  
  
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
475 TTTTGAACCAACAGCGCCGTATTGTGATGTAATGTGACTGAGATTGATG 524  
  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
::||| |||||

525 GTAAAGCAATGATTCGTATCCCTCGTTATGTCGACTTTCCTATTA 574  
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67  
|||||  
575 CCTGGACTACACTTACAAAGAAATAATGAATACTATGTGCAATGGCC 624  
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84  
|||||  
625 ATTAGATCGCAGACATATATAAGAGTTAGAGTAGTTGAATTAGATCCAA 674  
84 eraLalysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100  
|||||  
675 GCGCAAGATCGAAGTCACCTATATGATAGAATAGAAAAAGAGAGAA 724  
101 ThrLysSerPheProIleThrGlyLysGlyPheValValProAspLeuSe 117  
|||||  
725 ACGAGCTCTTCCCTATACAGAAAAAGGTTTGTGTGCCAGATTATAC 774  
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134  
|||||  
775 AGAGCATATTAAAAACCCTGATTCACCTAATTAACAAGGTTGTATAG 824  
134 LuLysLys 136  
|||||  
825 AAAAGAAA 832  
  
seq\_name: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-08-852-299-16  
  
seq\_documentation\_block:  
; Sequence 16, Application US/08852299  
; Patent No. 6010897  
;; GENERAL INFORMATION:  
;; APPLICANT: Behnke, Detlef  
;; APPLICANT: Schlott, Bernhard  
;; APPLICANT: Albrecht, Sybille  
;; APPLICANT: G hrs, Karl-Heinz  
;; APPLICANT: Hartmann, Manfred  
;; TITLE OF INVENTION: Expression of signal-peptide-free  
;; NUMBER OF SEQUENCES: 40  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Fish & Neave  
;; STREET: 1251 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: USA  
;; ZIP: 10020  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/852,299  
;; FILING DATE: 17-MAY-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/256,261  
;; FILING DATE:  
;; INFORMATION FOR SEQ ID NO: 16:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1023 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; ORIGINAL SOURCE:  
;; ORGANISM: Staphylococcus aureus  
;; STRAIN: phage 42D  
;; IMMEDIATE SOURCE:  
;; CLONE: Plasmid DB17  
;; FEATURE:  
;; NAME/KEY: -35\_signal



```
LOCATION: 224..229
FEATURE:
NAME/KEY: -10-signal
LOCATION: 248..253
FEATURE:
NAME/KEY: -
LOCATION: 332..336
OTHER INFORMATION: /note= "Shine-Dalgarno sequence"
FEATURE:
NAME/KEY: sig-peptide
LOCATION: 344..415
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 416..832
OTHER INFORMATION: /product= "mature protein SAK42D"
FEATURE:
NAME/KEY: CDS
LOCATION: 344..835
US-08-852-299-16
```

```
alignment_scores:
Quality: 695.00 Length: 136
Ratio: 5.148 Gaps: 0
Percent Similarity: 99.265 Percent Identity: 97.794
```

## alignment\_block:

US-09-728-670-10 x US-08-852-299-16 ..

Align seg 1/1 to: US-08-852-299-16 from: 1 to: 1023

```
1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTy 17
|||||
425 TCAAGTTCATTCGACAAAGAAATATATAAAGCGCATGACGAGTTA 474
|||||
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
475 TTTTGAACCAACAGCGCCGTTATTTGATGTAATGTGACTGAGTTGATG 524
|||||
34 erLysGlyAsnGluLeuSerProHisTyrValGluPheProIleLys 50
::|||
525 GTAAAGAAATGAATGCTATCCCTGTTATGTCGAGTTCCATTA 574
|||||
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTPrAl 67
|||||
575 CCTGGAGTACTACTTACAAAGAAATTAATACTATGTGCAATGGGC 624
|||||
67 aleuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84
|||||
625 ATTAGATCGCAGACATATTAAGAGTTAGAGTGAATTGATCCAA 674
|||||
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysLysGlu 100
|||||
675 GCGCAAGATCGAAGTCACTTATTATGATTAAGATAAGAAAAAGAGAA 724
|||||
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
|||||
725 ACGAAGTCTTCCCTATTAACAGAAAAAGTTTGTGTCACAGATTATC 774
|||||
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIle 134
|||||
775 AGAGCATATTAATAAACCTGATTCACACTTAATTACAAGGTTGTATAG 824
|||||
134 luLysLys 136
|||||
825 AAAAGAA 832
```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-256-261-11

## seq\_documentation\_block:

```
; Sequence 11, Application US/08256261
; Patent No. 5801037
; GENERAL INFORMATION:
```

```
APPLICANT: Behnke, Detlef
APPLICANT: Schlot, Bernhard
APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Helz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal-peptide-free
TITLE OF INVENTION: staphylokinases
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,261
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "recombinant DNA"
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Phage 42D
IMMEDIATE SOURCE:
CLONE: plasmid pMET5
FEATURE:
NAME/KEY: CDS
LOCATION: 1..414
FEATURE:
NAME/KEY: -
LOCATION: 1..3
OTHER INFORMATION: /note= "start codon"
FEATURE:
NAME/KEY: -
LOCATION: 412..414
OTHER INFORMATION: /note= "stop codon"
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 4..411
OTHER INFORMATION: /product= "mature protein"
US-08-256-261-11
```

## alignment\_scores:

```
Quality: 689.00 Length: 136
Ratio: 5.142 Gaps: 0
Percent Similarity: 98.529 Percent Identity: 97.059
```

## alignment\_block:

US-09-728-670-10 x US-08-256-261-11 ..

Align seg 1/1 to: US-08-256-261-11 from: 1 to: 414

```
1 SerSerSerPheAspLysGlyLysTyrLysGlyAspAspAlaSerTy 17
|||||
4 TCAAGTTCATTCGACAAAGAAATATATAAAGCGCATGACGAGTTA 53
|||||
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||
54 TTTTGAACCAACAGCGCCGTTATTTGTGTAATGTGACTGAGTYGAYG 103
|||||
```

```
34 erlysglyAsnGluLeuSerProHisTyrValGluPheProIleLys 50
:|||||
104 GTAAAGAAATGARGTSTYRTRCCCKGKTATGTCAGAGTTTCTATTAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrpAl 67
|||||
154 CCTGGGACTACACTTACAAAGAAAAATTCGAATCTATGTGCAATGGGC 203
67 aLeuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspProS 84
|||||
204 ATTAGATCGCAGCAGCATATTAAGAGTTTAGAGTAGTTGAATTAGATCCAA 253
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
254 GCGCAAGATCGAAGTCAGTATTATGATAGAATAGAAAAAGAGAA 303
304 ACGAAGTCTTTCCCTATACAGAAAAAGGTTTGTGTCCAGATTTATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
354 AGAGCATATTAATAAACCTGATTCACCTTAATTACCAAGGTGTATAG 403
134 LuLysLys 136
|||||
404 AAAAGAAA 411
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-852-299-11
seq_documentation_block:
; Sequence 11, Application US/08852299
; Patent No. 6010897
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlott, Bernhard
; APPLICANT: Albrecht, Sybille
; APPLICANT: G hrs, Karl-Helz
; APPLICANT: Hartmann, Manfred
; TITLE OF INVENTION: Expression of signal-peptide-free
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,299
; FILING DATE: 17-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/256,261
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "recombinant DNA"
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Phage 42D
; IMMEDIATE SOURCE:
```

```
CLONE: plasmid pMET5
FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..414
FEATURE:
; NAME/KEY: -
; LOCATION: 1..3
OTHER INFORMATION: /note= "start codon"
FEATURE:
; NAME/KEY: -
; LOCATION: 412..414
OTHER INFORMATION: /note= "stop codon"
FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4..411
OTHER INFORMATION: /product= "mature protein"
US-08-852-299-11
alignment_scores:
Quality: 689.00 Length: 136
Ratio: 5.142 Gaps: 0
Percent Similarity: 98.529 Percent Identity: 97.059
alignment_block:
US-09-728-670-10 x US-08-852-299-11 ..
Align seg 1/1 to: US-08-852-299-11 from: 1 to: 414
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
|||||
4 TCAGAGTCATTCGACAAAGGAAATATTAAGAGCGCATGACGCGAGTTA 53
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspS 34
|||||
54 TTTTGAACCAACAGCGCCGCTATTGTGTGTAATGTGACTGCACTGAYG 103
34 erlysglyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
:|||||
104 GTAAAGAAATGARGTSTYRTRCCCKGKTATGTCAGAGTTTCTATTAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrrpAl 67
|||||
154 CCTGGGACTACACTTACAAAGAAAAATTCGAATCTATGTGCAATGGGC 203
67 aLeuAspAlaThrAlaTyrLysGluPheArgValGluLeuAspProS 84
|||||
204 ATTAGATCGCAGCAGCATATTAAGAGTTTAGAGTAGTTGAATTAGATCCAA 253
84 eAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
|||||
254 GCGCAAGATCGAAGTCAGTATTATGATAGAATAGAAAAAGAGAA 303
304 ACGAAGTCTTTCCCTATACAGAAAAAGGTTTGTGTGTCCAGATTTATC 353
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSe 117
|||||
304 ACGAAGTCTTTCCCTATACAGAAAAAGGTTTGTGTGTCCAGATTTATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleG 134
|||||
354 AGAGCATATTAATAAACCTGATTCACCTTAATTACCAAGGTGTATAG 403
134 LuLysLys 136
|||||
404 AAAAGAAA 411
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-256-261-13
seq_documentation_block:
; Sequence 13, Application US/08256261
; Patent No. 5801037
; GENERAL INFORMATION:
; APPLICANT: Behnke, Detlef
; APPLICANT: Schlott, Bernhard
```

```

: APPLICANT: Albrecht, Sybille
: APPLICANT: G hrs, Karl-Heinz
: APPLICANT: Hartmann, Manfred
: TITLE OF INVENTION: Expression of signal-peptide-free
: TITLE OF INVENTION: staphylokinases
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Neave
: STREET: 1251 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10020
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/256,261
: FILING DATE:
: CLASSIFICATION: 435
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 414 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "recombinant DNA"
: ORIGINAL SOURCE:
: ORGANISM: Staphylococcus aureus
: STRAIN: phage 42D
: IMMEDIATE SOURCE:
: CLONE: Plasmid pMET5
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 4..411
: OTHER INFORMATION: /product= "mature protein"
: FEATURE:
: NAME/KEY: -
: LOCATION: 1..3
: OTHER INFORMATION: /note= "start codon"
: FEATURE:
: NAME/KEY: -
: LOCATION: 412..414
: OTHER INFORMATION: /note= "stop codon"
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..414
: US-08-256-261-13

alignment_scores:
  Quality: 687.00      Length: 136
  Ratio: 5.127        Gaps: 0
  Percent Similarity: 98.529  Percent Identity: 96.324

alignment_block:
US-09-728-670-10 x US-08-256-261-13 ..
Align seg 1/1 to: US-08-256-261-13 from: 1 to: 414

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17
  |||||||
4 TCAAGTTCATTCGACAAAGGAAATATAAAAAAGCGATGACCGAGTTA 53
  |||||||
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
  |||||||
54 TTTTGAACCAACAGGCCCGTATTGCTAGTAATGTGACTGAGGTGAYG 103
  |||||||
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50
  ::||| |||||||
```

```

104 GTAAAGAAATGARCSTYTRTCCCKCKRTATGTCAGTTTCCTATTAA 153
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAl 67
  |||||||
154 CCTGGACTACACTTACAAAAAGAAAAATGAACTATGTCGAATGGGC 203
67 aleuAspAlaThrAlaTyrLysGluPheArgValAlaGluLeuAspPro 84
  |||||||
204 ATTAGATGCAGACATATAAAGATTAGAGTAGTGAATTAGATCCAA 253
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGlu 100
  |||||||
254 GCGCAAGATCGAGTCACCTATTATGATAGATAAGAAAAAGAGAA 303
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117
  |||||||
304 ACGAAGTCTTCCCTATAACAGAAAAAGGTTTGTGTCACAGATTATC 353
117 rGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValIleG 134
  |||||||
354 AGAGCATATTAAAAACCTCGATTCAACTTAATTACMAAGTTGTTATAG 403
134 luLysLys 136
  |||||||
404 AAAAGAAA 411
```

seq\_name: /cgn2\_6/ptodata/1/lna/6A\_COMB.seq:US-08-852-299-13

seq\_documentation\_block:

; Sequence 13, Application US/08852299  
; Patent No. 6010897

; GENERAL INFORMATION:

; APPLICANT: Behnke, Detlef  
; APPLICANT: Schlotz, Bernhard  
; APPLICANT: Albrecht, Sybille  
; APPLICANT: G hrs, Karl-Heinz  
; APPLICANT: Hartmann, Manfred  
; TITLE OF INVENTION: Expression of signal-peptide-free  
; TITLE OF INVENTION: staphylokinases  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,299  
; FILING DATE: 17-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/256,261  
; FILING DATE:

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 414 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "recombinant DNA"

; ORIGINAL SOURCE:

; ORGANISM: Staphylococcus aureus

; STRAIN: phage 42D

; IMMEDIATE SOURCE:

; CLONE: Plasmid pMET5

; FEATURE:

NAME/KEY: mat\_peptide  
LOCATION: 4..411  
OTHER INFORMATION: /product= "mature protein"  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..3  
OTHER INFORMATION: /note= "start codon"  
FEATURE:  
NAME/KEY: -  
LOCATION: 412..414  
OTHER INFORMATION: /note= "stop codon"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..414  
US-08-852-299-13

alignment\_scores:  
Quality: 687.00 Length: 136  
Ratio: 5.127 Gaps: 0  
Percent Similarity: 98.529 Percent Identity: 96.324

alignment\_block:  
US-09-728-670-10 x US-08-852-299-13 ..

Align seg 1/1 to: US-08-852-299-13 from: 1 to: 414

1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAspAlaSerTy 17  
|||||  
4 TCAAGTTCATTCGACAAAGAAATATAAAAAGCGATGACGCGAGTTA 53  
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34  
|||||  
54 TTTTGAACCAACACAGCGCCGCTATTGCTAGTAATGCTGAGCTGAGYTGA 103  
34 erLysGlyAsnGluLeuLeuSerProHisTyrValGluPheProIleLys 50  
:::|||||  
104 GTAAAAGAAATGARCTSTRTCCTCCCKGKTATGTCGAGTTCTCTATTAA 153  
51 ProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPal 67  
|||||  
154 CTTGGGACTACACTTACAAAAGAAAAATTGAATTAATGTCGAATGGGC 203  
67 aLeuAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspPro 84  
|||||  
204 ATTAGATGCGACAGCATATTAAGAGTTAGAGTAGTTGAATTAGATCCAA 253  
84 erAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluLys 100  
|||||  
254 GCGCAAAAGATCGAAGTCACTTATTATGATAGAAATGAAGAAAAAGAGAA 303  
101 ThrLysSerPheProIleThrGluLysGlyPheValValProAspLeu 117  
|||||  
304 ACGAAGTCTTCCCTATACAGAAAAAGTTTGTGTCACAGATTATATC 353  
117 rGluHisIleLeuAsnProGlyPheAsnLeuIleThrLysValValIle 134  
|||||  
354 AGAGCATATTAATAAACCTGATTCACACTTAATTACMAAGTTGTATTAG 403  
134 LuLysLys 136  
|||||  
404 AAAAGAAA 411

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-256-261-7

seq\_documentation\_block:

; Sequence 7, Application US/08256261  
; Patent No. 5801037  
; GENERAL INFORMATION:  
; APPLICANT: Behnke, Detlef  
; APPLICANT: Schlot, Bernhard  
; APPLICANT: Albrecht, Sybille  
; APPLICANT: G hrs, Karl-Heinz

APPLICANT: Hartmann, Manfred  
TITLE OF INVENTION: Expression of signal-peptide-free  
TITLE OF INVENTION: staphylokinases  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Neave  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/256,261  
FILING DATE:  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 384 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "recombinant DNA"  
ORIGINAL SOURCE:  
ORGANISM: Staphylococcus aureus  
STRAIN: phage 42D  
IMMEDIATE SOURCE:  
CLONE: Plasmid pMET5  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..3  
OTHER INFORMATION: /note= "start codon"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..384  
FEATURE:  
NAME/KEY: -  
LOCATION: 382..384  
OTHER INFORMATION: /note= "stop codon"  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 4..381  
OTHER INFORMATION: /product= "shortened mature  
OTHER INFORMATION: protein"  
US-08-256-261-7

alignment\_scores:  
Quality: 643.00 Length: 126  
Ratio: 5.144 Gaps: 0  
Percent Similarity: 99.206 Percent Identity: 97.619

alignment\_block:  
US-09-728-670-10 x US-08-256-261-7 ..

Align seg 1/1 to: US-08-256-261-7 from: 1 to: 384

11 LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVa 27  
|||||  
4 AAAGCGCATGACCGCGAGTTATTGTAACCAACAGCGCCGTTATTGATGCT 53  
27 LasnValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHis 44  
|||||  
54 AAATGTCAGTGAAGTTGATGTAAGAAGAAATGAATGCTATCCCTCGTT 103  
44 yrValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIle 60  
|||||  
104 ATGTGAGTTCTCTATTAAACCTGGGACTACACTTACAAAGAAAAAATTT 153



```
61  GUTYRTRValGUTrPalLeuAspAlaThrAlaTyrLysGluPheAr 77
      |||
154  GAATAGTACTATGTCGAATGGCATTAGATGCGACACATATAAGAGTTTAG 203
      |||
77  gValValGluLeuAspProSerAlaLysIleGluValThrTyrTrAspL 94
      |||
204  AGTAGTTGATGATGATCCAGCGCAAGATCGAAGTCACCTTATATGATA 253
      |||
94  ySAsnLysLysGluGluThrLysSerPheProIleThrGluLysGly 110
      |||
254  AGAATTAAGAAAAAGAGAAACGAAGCTTTCCCTATAACAGAAAAAGGT 303
      |||
111  PheValValProAspLeuSerGluHisIleLysAsnProGlyPheAsnL 127
      |||
304  TTTGTTGTCAGATTTATACAGACATATTAAAAACCTGATTCAACTT 353
      |||
127  uIleThrLysValValIleGluLysLys 136
      |||
354  AATTACMAAGGTTGTTATAGAAAAAGAAA 381
```

seq\_name: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:US-08-852-299-7

seq\_documentation\_block:

; Sequence 7, Application US/08852299

; Patent No. 6010897

; GENERAL INFORMATION:

; APPLICANT: Behnke, Detlef

; APPLICANT: Schloft, Bernhard

; APPLICANT: Albrecht, Sybille

; APPLICANT: G hrs, Karl-Heinz

; APPLICANT: Hartmann, Manfred

; TITLE OF INVENTION: Expression of signal-peptide-free

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Neave

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/852,299

; FILING DATE: 17-MAY-1997

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 08/256,261

; FILING DATE:

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 384 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "recombinant DNA"

; ORIGINAL SOURCE:

; ORGANISM: Staphylococcus aureus

; STRAIN: phage 42D

; IMMEDIATE SOURCE:

; CLONE: Plasmid pMET5

; FEATURE:

; NAME/KEY: -

; LOCATION: 1..3

; OTHER INFORMATION: /note= "start codon"

; FEATURE:

; NAME/KEY: CDS

```
; LOCATION: 1..384
; FEATURE:
; NAME/KEY: -
; LOCATION: 382..384
; OTHER INFORMATION: /note= "stop codon"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 4..381
; OTHER INFORMATION: /product= "shortened mature
; OTHER INFORMATION: protein"
; US-08-852-299-7
```

```
alignment_scores:
  Quality: 643.00      Length: 126
  Ratio: 5.144        Gaps: 0
Percent Similarity: 99.206  Percent Identity: 97.619
```

alignment\_block:

US-09-728-670-10 x US-08-852-299-7 ..

Align seg 1/1 to: US-08-852-299-7 from: 1 to: 384

```
11  LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVa 27
      |||
4  AAAGCGATGACGCGAGTTATTGTAACCAACAGAGCCCGTATTGATGCT 53
      |||
27  lAsnValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHist 44
      |||
54  AAATGTACTGACGAGTTGATGTAAAGAAATGAATTGCTATCCCTCGTT 103
      |||
44  yValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIle 60
      |||
104  ATGTCAGATTTCCTATTAAACCTGGACTACACTTACAAAGAAAAAATT 153
      |||
61  GUTYRTRValGUTrPalLeuAspAlaThrAlaTyrLysGluPheAr 77
      |||
154  GAATAGTACTATGTCGAATGGCATTAGATGCGACACATATAAGAGTTTAG 203
      |||
77  gValValGluLeuAspProSerAlaLysIleGluValThrTyrTrAspL 94
      |||
204  AGTAGTTGATGATGATCCAGCGCAAGATCGAAGTCACCTTATATGATA 253
      |||
94  ySAsnLysLysGluGluThrLysSerPheProIleThrGluLysGly 110
      |||
254  AGAATAAGAAAAAGAGAAACGAAGCTTTCCCTATAACAGAAAAAGGT 303
      |||
111  PheValValProAspLeuSerGluHisIleLysAsnProGlyPheAsnL 127
      |||
304  TTTGTTGTCAGATTTATACAGACATATTAAAAACCTGATTCAACTT 353
      |||
127  uIleThrLysValValIleGluLysLys 136
      |||
354  AATTACMAAGGTTGTTATAGAAAAAGAAA 381
```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-256-261-9

seq\_documentation\_block:

; Sequence 9, Application US/08256261

; Patent No. 5801037

; GENERAL INFORMATION:

; APPLICANT: Behnke, Detlef

; APPLICANT: Schloft, Bernhard

; APPLICANT: Albrecht, Sybille

; APPLICANT: G hrs, Karl-Heinz

; APPLICANT: Hartmann, Manfred

; TITLE OF INVENTION: Expression of signal-peptide-free

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Neave

; STREET: 1251 Avenue of the Americas

; CITY: New York

```
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,261
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "recombinant DNA"
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: phage 42D
IMMEDIATE SOURCE:
CLONE: plasmid pMET5
FEATURE:
NAME/KEY: -
LOCATION: 1..3
OTHER INFORMATION: /note= "start codon"
FEATURE:
NAME/KEY: CDS
LOCATION: 1..372
FEATURE:
NAME/KEY: -
LOCATION: 370..372
OTHER INFORMATION: /note= "stop codon"
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 4..369
OTHER INFORMATION: /product= "shortened mature
OTHER INFORMATION: protein"
US-08-256-261-9

alignment_scores:
Quality: 620.00 Length: 122
Ratio: 5.124 Gaps: 0
Percent Similarity: 99.180 Percent Identity: 97.541

alignment_block:
US-09-728-670-10 x US-08-256-261-9 ..
Align seg 1/1 to: US-08-256-261-9 from: 1 to: 372

15 Alasertyrphggluprothrglyprotyrleumetvalasnavalthrcl 31
|||||
4 GCGAGTTATTGTAACCAACAGCCCGTATTGATGTTAAATGTGACTGG 53
|||||
31 yvalaspsertysglyasngluleuleuserprohistytrvalglupher 48
|||||
54 AGTTGATGGTAAAGAAATGAAATGCTATCCCTCGTTATGTGAGTTTC 103
|||||
48 rollelysproglythrthrleuthrlysglulyslleglutyrttyrval 64
|||||
104 CTATTAAACCTGGGACTACACTTACAAGAAAAAATTGAATACTATGTC 153
|||||
65 GluTrpAlaLeuAspAlaThrAlaTyrTyrGluPheArgValAlaGlu 81
|||||
154 GAATGGGCATTAGATGCGACAGCATATAAGAGTTTAGAGTAGTGAATT 203
|||||
81 uaspproseralalysilegluvalthrtyrtyraspLysasnLysL 98
|||||
204 AGATCCAAAGCGCAAGATCGAAGTCACTTATATGATAGAATAAGAAAA 253
|||||
```

```
98 ysglglutThrLysSerPheProIleThrGluLysGlyPheValValPro 114
|||||
254 AAGAAGAAACGAAGTCTTCCCTATTAACAGAAAGCTTTGTGTCCCA 303
|||||
115 AspleuserGluHisIleLysAsnProGlyPheAsnLeuIleThrLysVa 131
|||||
304 GATTATACAGACATATTAATAAACCCGTGATTCACCTTATTACMAAGGT 353
|||||
131 lvalileglulyslys 136
|||||
354 TGTTATAGAAAAAGAA 369
|||||

seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-852-299-9
seq_documentation_block:
Sequence 9, Application US/08852299
Patent No. 6010897
GENERAL INFORMATION:
APPLICANT: Behnke, Detlef
APPLICANT: Schlott, Bernhard
APPLICANT: Albrecht, Sybille
APPLICANT: G hrs, Karl-Heinz
APPLICANT: Hartmann, Manfred
TITLE OF INVENTION: Expression of signal-peptide-free
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,299
FILING DATE: 17-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/256,261
FILING DATE:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "recombinant DNA"
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Phage 42D
IMMEDIATE SOURCE:
CLONE: plasmid pMET5
FEATURE:
NAME/KEY: -
LOCATION: 1..3
OTHER INFORMATION: /note= "start codon"
FEATURE:
NAME/KEY: CDS
LOCATION: 1..372
FEATURE:
NAME/KEY: -
LOCATION: 370..372
OTHER INFORMATION: /note= "stop codon"
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 4..369
```

OTHER INFORMATION: /product= "shortened mature  
OTHER INFORMATION: protein"  
US-08-852-299-9

alignment\_scores:  
Quality: 620.00 Length: 122  
Ratio: 5.124 Gaps: 0  
Percent Similarity: 99.180 Percent Identity: 97.541

alignment\_block:  
US-09-728-670-10 x US-08-852-299-9 ..

Align seg 1/1 to: US-08-852-299-9 from: 1 to: 372

```
15 AlaSerTyrPheGluProThrGlyProTyrLeuMetValAsnValThrG1 31
|||||
4 GCGAGTATTATTGAACCAACAGCCCGTATTGATGGTAATGTGACTGG 53
31 yValAspSerLysGlyAsnGluLeuSerProHisTyrValGluPhe 48
|||||
54 AGTGTAGTAAAGAAATGAATGCTATCCCTCGTATGTGCGAGTTTC 103
48 rolleThrGlyThrThrLeuThrLysGluLysIleGluTyrVal 64
|||||
104 CTATTAACCTGGGACTACACTTACAAAGAAAAAATGAATACTATGTC 153
65 GluTrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValGluLe 81
|||||
154 GAATGGGCATTAGATCGCAGACAGCATATAAGAGTTAGAGTAGTGAAT 203
81 uAspProSerAlaLysIleGluValThrTyrTyrAspLysAsnLysL 98
|||||
204 AGATCCAGCCGCAAGATCGAGTCACTTATTATGATAAGATAAGAAA 253
98 ysgGluGluThrLysSerPheProIleThrGluLysGluPheValP 114
|||||
254 AAGAAGAAACGAAGCTTTCCTTAATACAGAAAAAGTTTGTGTCCCA 303
115 AspLeuSerGluHisIleLysAsnProGlyPheAsnLeuIleThrL 131
|||||
304 GATTATTCAGAGCATATTAAACCCTGATTCAACTTAATTACMAAGGT 353
131 lValIleGluLysLys 136
|||||
354 TGTATAGAAAAAGAAA 369
```

seq\_name: /cgn2\_6/ptodata/1/lna/backfiles1.seq:5182210-9

seq\_documentation\_block:

Patent No. 5182210  
APPLICANT: BINNS, MATTHEW M.;BOURSNELL, MICHAEL E.G.;  
CAMPBELL, JOAN I.A.;TOMLEY, FIONA M.  
TITLE OF INVENTION: FOWLPOX VIRUS PROMOTERS  
NUMBER OF SEQUENCES: 22  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,608  
FILING DATE: 21-OCT-1988  
SEQ ID NO:9:  
LENGTH: 11225  
5182210-9

alignment\_scores:

Quality: 84.50 Length: 123  
Ratio: 1.408 Gaps: 5  
Percent Similarity: 48.780 Percent Identity: 25.203

alignment\_block:

US-09-728-670-10 x 5182210-9/rev ..

Align seg 1/1 to reverse of: 5182210-9 from: 1 to: 11225

```
43 HisTyrValGluPheProIleLysProGlyThrThrLeuThrLysGlu 59
|||||
1113 CATATTGTTGAACCTATACATATAGCTGCTAATAGAACAGAAAGTAA 11064
59 sile..GluTyrTyrValGluTrpAlaLeuAspAlaThrAla..... 72
|||||
11063 AATAGTAAATGCTTATAGATACGGCGCTGACATTAAATTCAGAGACG 11014
73 .....TyrLysGlu 75
11013 GCGCGAATGGTAAATACCCCTATACATTACGCTATGAAAGTATAGATCCG 10964
76 PheArgVal.....ValGluLeuAspProSerAlaLysIleG1 88
|||||
10963 TTTAGATTAAATAATAAAGTATTATTAGACCACGGCGCC..... 10922
88 uValThrTyrTyrAspLysAsnLysLysGluGluThrLysSerPhe 105
|||||
10921 .....GATATTACAAACAAACCGCTTTTAACATAATACATCCC 10885
105 rolleThrGluLysGlyPheValValProAspLeuSerGluHisIle 121
|||||
10884 CCTTATACGAACCTAGGTTTATTACCAGCAGCCTTATTAGATTACATCA 10835
122 AsnProGlyPheAsnLeu.....IleTh 129
|||||
10834 TCTAGAGAGCTAATATATAATAAAGAAAGATGGGTAGATATATTAC 10785
129 rLysValValIleGluLys 135
|||||
10784 ACGAATAATATTAGACGA 10766
```

seq\_name: /cgn2\_6/ptodata/1/lna/6B\_COMB.seq:US-09-321-831-1

seq\_documentation\_block:

Sequence 1, Application US/09321831  
Patent No. 6262336  
GENERAL INFORMATION:  
APPLICANT: LUBON, Henryk  
APPLICANT: DROHAN, William  
APPLICANT: HENNIGHAUSEN, Lothar  
APPLICANT: VELANDER, William  
TITLE OF INVENTION: Expression of a Heterologous Protein C in Mammary Tissue of  
FILE REFERENCE: 030523/0150  
CURRENT APPLICATION NUMBER: US/09/321,831  
PRIOR FILING DATE: 1999-05-28  
PRIOR APPLICATION NUMBER: US 09/184,163  
PRIOR FILING DATE: 1998-11-02  
PRIOR APPLICATION NUMBER: US 07/943,246  
PRIOR FILING DATE: 1992-09-10  
PRIOR APPLICATION NUMBER: US 07/638,995  
PRIOR FILING DATE: 1991-01-11  
PRIOR APPLICATION NUMBER: PCT/US98/02638  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 08/443,184  
PRIOR FILING DATE: 1995-05-17  
PRIOR APPLICATION NUMBER: US 08/198,068  
PRIOR FILING DATE: 1994-02-18  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 4122  
TYPE: DNA  
ORGANISM: WAP gene promoter fragment from the C57B/6 mouse strain  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (544)..(544)  
OTHER INFORMATION: "n" is unknown  
NAME/KEY: misc\_feature  
LOCATION: (3377)..(544)  
OTHER INFORMATION: "n" is unknown  
US-09-321-831-1

alignment\_scores:                      length:        115  
                    Quality:        81.50  
                    Ratio:        1.509        Gaps:        6  
Percent Similarity: 46.957        Percent Identity: 29.565

## alignment\_block:

US-09-728-670-10 x US-09-321-831-1/rev ..

Align seg 1/1 to reverse of: US-09-321-831-1 from: 1 to: 4122

```
42 ProHISTyValGlupheProIleLysProGlyThrThrLeuThrLysG1 58
   ||||| :::: ||| ||||| ||| ||
3981 CCACACAGAGAGAGAGCCCGAGCTGCCAGAACAAAGGCCACCAATGGA 3932
58 ULysIleGluTyrTyrValGluTrpAlaLeuAspAlaThrAlaTyrLysG 75
   ||| ||| :::: |||
3931 GCCCATTT.....TTTAAA 3918
75 Lu.....PheArgValValGluLeuAspProSerAlaLysIleGluVal 89
   :: ||||| :::: ||| :::: |||
3917 AGTCCTTGTAGGGTGTCAATTTTCCCCCACTTAAAAAGTAATACTT 3868
90 ThrTyrTyrAspLys.AsnLysLysLysGluGluThrLysSer..... 103
   ::||| :::: ||| :::: |||
3867 GTTATTATGAAGCGGAGAAATGAAAGAGACGAGAAAAAGAAATAGG 3818
104 .....PhePro 105
3817 ATCAGTGACAAATGTGACAACTGTCACTTGTGCTGGTGGAGACC 3768
106 IleThrGluLysGly..PheValValProAspLeuSerGluHisIleLys 121
   ||||| :::: ||| :::: |||
3767 ATCACAGACAAAGCAATCTGAGCTCCAAATTTGTACAGTCACATAGAA 3718
```

```
122 Asn.....ProGlyPhe 125
   ||| ||| :::: |||
3717 AACAAAGACAGAGTTTGAAGCAAGCTCGAGTCGGGCTAC 3676
```

seq\_name: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq:US-08-525-654A-136

## seq\_documentation\_block:

; Sequence 136, Application US/08525654A

; Patent No. 5736356

## ; GENERAL INFORMATION:

; APPLICANT: SANO, KOHICHIRO

; APPLICANT: KUMAZAWA, YOSHIYUKI

; APPLICANT: YASEUDA, HISASHI

; APPLICANT: SEGURO, KATSUYA

; APPLICANT: MOTOKI, MASAO

; TITLE OF INVENTION: TRANSGlutAMINASE ORIGINATED FROM

; NUMBER OF SEQUENCES: 150

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &amp; NEUSTADT,

; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

; CITY: ARLINGTON

; STATE: VA

; COUNTRY: USA

; ZIP: 22202

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; FILING DATE: 28-SEP-1995

; APPLICATION NUMBER: US/08/525,654A

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: JP 6/8283

; FILING DATE: 28-JAN-1994

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: JP 7/3876

; FILING DATE: 13-JAN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: OBLON, NORMAN F.

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 10-760-0 PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-413-3000

; TELEFAX: 703-413-2220

; INFORMATION FOR SEQ ID NO: 136:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3394 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ORIGINAL SOURCE:

; ORGANISM: Crassostrea gigas

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 305..2617

US-08-525-654A-136

alignment\_scores:                      length:        90  
                    Quality:        80.00  
                    Ratio:        1.379        Gaps:        5  
Percent Similarity: 64.444        Percent Identity: 26.667

## alignment\_block:

US-09-728-670-10 x US-08-525-654A-136 ..

Align seg 1/1 to: US-08-525-654A-136 from: 1 to: 3394

```
25 LeuMetValAsnValThrGlyValAspSerLysGlyAsnGluLeuLeu 41
   ||| :::: ||| :::: |||
2369 CTGCTGCTACCCAGGCTGAATAGCCGTGAGGGCTCGGCTCTCTAGC 2418
41 rProHISTyValGluPhe.....ProIleLysProGly..... 52
   ||| :::: ||| :::: |||
2419 TCCACAACCATAGATATAGACAGCTCTATTAAACCTGTGACGAGGTCA 2468
53 ....ThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAla 67
   ||| :::: ||| :::: |||
2469 AGAAACAGTCGCTCTCGCCCCCGCAACCATACTAC.....TGGGGT 2512
68 LeuAspAlaThrAla..TyrLysGluPheArgValValGluLeuAspR 83
   :: ||| :::: ||| :::: |||
2513 CGTGAAGTATAGCGACCTTCACTCCAAACAGATGTGACATTGAGAC 2562
83 OSerAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluG 100
   ||||| ||| :::: ||| :::: |||
2563 CAGCGCGCATTAAGTGTATCGACACAGAACAAAGATAGACAGCGACA 2612
100 Lu.ThrLysSerPhePro 105
   :: ||| :::: |||
2613 GTGACTAGAGATTCCTCA 2630
```

seq\_name: /cgn2\_6/ptodata/1/lna/6B\_COMB.seq:US-09-251-645-11

## seq\_documentation\_block:

; Sequence 11, Application US/09251645

; Patent No. 6281413

## ; GENERAL INFORMATION:

; APPLICANT: Kramer, Vance C.

; APPLICANT: Morgan, Michael K.

; APPLICANT: Anderson, Arne R.

; APPLICANT: Hart, Hope

; APPLICANT: Warren, Gregory W.

; APPLICANT: Dunn, Martha

; APPLICANT: Chen, Jeng S.



; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS  
; FILE REFERENCE: CGC1963/A  
; CURRENT APPLICATION NUMBER: US/09/251,645  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 37948  
; TYPE: DNA  
; ORGANISM: Photorhabdus luminescens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (15171)..(18035)  
; OTHER INFORMATION: orf5  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (23768)..(31336)  
; OTHER INFORMATION: hph2  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (31393)..(35838)  
; OTHER INFORMATION: orf2  
; US-09-251-645-11

alignment\_scores:  
Quality: 73.50 Length: 128  
Ratio: 1.148 Gaps: 8  
Percent Similarity: 50.000 Percent Identity: 28.906

alignment\_block:  
US-09-728-670-10 x US-09-251-645-11 ..

Align seg 1/1 to: US-09-251-645-11 from: 1 to: 37948

12 Glyaspasp.....Alasertyrphggluprothrglyprotyrle 25  
||||||| ||| ::|||::  
28148 GGTGATGATATGCTACAGACTCGTTAAACAACCGAATGATCTTAAGCA 28197  
25 uMetValasnValThrGlyValaspSerlysgly..... 36  
||| ::||| ||||| |||||  
28198 ATACGTCTATATGACT.....GACAGTAAAGGTACTGCTACCGATGCT 28241  
37 .....AsngluleuLeuSerProHisTyrValgluphe 47  
||| ::||| ||||| |||||  
28242 CAGACCAAGTAGATATCAATGCTGCAATTTCCCGCAAAAGTTCAGGTA 28291  
48 ProtleysProglyThr.....ThrLeuThrlysglulys..11 60  
::||| ||||| ::|||:: |||||::  
28292 ACAGTAAAGCCGCGTAGCAAGAACAAACGTTTACCGCGGATAAAAATGT 28341  
60 egluTyrTyrValgluTrrPalaleuaspAlaThrAlaTyrlysgluphea 77  
::||| ::||| ||||| |||||  
28342 CTCGATTCAGCCATCCCTAGCTTGTGATGAATGAATATAT...CAATTTA 28388  
77 rglValValgluleuasp.....ProSerAla 85  
::|||::|||::||| |||||  
28389 ATGCTCTCGAAATAGATGGCTCAAGTCTGAATTTTACTTAACAATTACGCC 28438  
86 LysIleGlulValThrTyr.....TyrAspLysasnLysLysly 98  
::|||::|||::||| |||||  
28439 AGTATGATATATTACCTTACCGCATTTTGACAGAGATGACGTAACCTGGG 28488  
98 sgluGlulThrLysSerPheProIleThrGlulys 109  
|||::|||::||| |||||  
28489 TTATGAAGTTTCAGTATTCCTATTACCGCAAG 28522

seq\_name: /cgn2\_6/ptodata/1/1na/5A\_COMB.seq:US-08-480-481-3

seq\_documentation\_block:  
; Sequence 3, Application US/08480481  
; Patent No. 5639614

; GENERAL INFORMATION:  
; APPLICANT: Sen, Luy1  
; APPLICANT: Phillipson, Kenneth D.  
; APPLICANT: Lusis, Aldons J.  
; TITLE OF INVENTION: GENE MUTATION IN PATIENTS WITH IDIOPATHIC DILATED CARD  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,481  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: UCLAO12.001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 714-760-0404  
; TELEFAX: 714-760-9502  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1050 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; FEATURE:  
; NAME/KEY: Other  
; LOCATION: 1...1  
; OTHER INFORMATION: Position 1 corresponds to  
; US-08-480-481-3  
; OTHER INFORMATION: 7987 of the rabbit cdna homolog

alignment\_scores:  
Quality: 73.00 Length: 108  
Ratio: 1.281 Gaps: 6  
Percent Similarity: 52.778 Percent Identity: 25.000

alignment\_block:  
US-09-728-670-10 x US-08-480-481-3 ..

Align seg 1/1 to: US-08-480-481-3 from: 1 to: 1050

17 TyrphgluProthrglyProtyrleuMetValasnValThrGlyValas 33  
|||::||| ::|||:: |||||::  
55 TACATGAATCAAAATTATGTCAGTATGATGAAAAACAGTCATCAATGCA 104  
33 pserlysglyasnGluleuLeuSerProHisTyrValglupheProIlel 50  
|||::|||::||| ::|||:: |||||::  
105 TTCTGAAGGCAAC.....TTTAACCCACAA.....CCGTGTG 136  
50 yspProglyThrThrleuThrLysGlulysIleGlulTyrTyrVal..... 64  
::|||::|||::||| |||||  
137 ATACCTCAAAATATATAATTCCTGAGAGTTGGAATCTCATTAACAAA 186

```

65 .....GlUTPAlaleuAspAlaThrAla..... 72
    ::|||:::||||
187 TATGCAGACATTCATGCACAAATGGTCATCATGACACAAGTTAGCAAAATGG 236
73 .....TyrlSegluPheargValValGluleuAspProSerAlaLysI 87
    ||| |||:::
237 ATGATTATATGAGAAATA.....TATTCAGACTCCTCCAGA 274
87 leglu.....ValThrTYrTYrasplysAsnLysLysLysGluiu 100
    ||::: ||| ::::|||:::||||
275 TTCAACCCCTTGATGAACAACCATATAAACTATATCTGAAAAAGAAAAAGAA 324
101 ThrLysSerPheProIleThrglu 108
    :::||||| |||
325 ATTATCGCTGGCCCAATCAAGAA 348

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-480-481-4

seq_documentation_block:
; Sequence 4, Application US/08480481
; Patent No. 5639614
; GENERAL INFORMATION:
; APPLICANT: Sen, Luyt
; APPLICANT: Phillips, Kenneth D.
; APPLICANT: Lusis, Aldons J.
; TITLE OF INVENTION: GENE MUTATION IN PATIENTS WITH IDIOPATHIC DILATED CARD
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,481
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: UCLA012.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1050 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Other
; LOCATION: 380..380
; OTHER INFORMATION: position distinguishes idiopathic
; OTHER INFORMATION: dilated cardiomyopathy
; NAME/KEY: other
; LOCATION: 776..776
;

```

```

; OTHER INFORMATION: Position distinguishes idiopathic
; OTHER INFORMATION: dilated cardiomyopathy
; NAME/KEY: Other
; LOCATION: 1...1
; OTHER INFORMATION: Position 1 corresponds to
; OTHER INFORMATION: 7987 of the rabbit cDNA homolog
US-08-480-481-4

alignment_scores:
    quality:      73.00          length:      108
    ratio:        1.281         gaps:         6
    percent similarity: 52.778   percent identity: 25.000

alignment_block:
US-09-728-670-10 x US-08-480-481-4 ..

Align seg 1/1 to: US-08-480-481-4 from: 1 to: 1050

17 TyrPheGluProThrglyProTyrLeuMetValasnValThrGlyValas 33
||||:||||| :: ||||| :::::|||||
55 TACATGGAATCAAAATTATGTCTAGTATGATGGAAAACAAGTCATCAATGGA 104
33 pSerLysGLYAsnGLUleuLeuSerProHisTYrVALgluPheProIlel 50
||||:||||| :::::||||| :|||:
105 TTCTGAAGGGAAC.....TTTAACCCACAA.....CCTGTTG 136

50 ySProGLyThrThrLeuThrLysGLULysILeGLUTyrTYrVal..... 64
::::: :: |||||:|||||:
137 ATACCTCAAATATTATAATTCTCGAAGAAGTTGGAATACTTCAATTAACAA 186

65 .....GlUTRpAlaLeuAspaLaThRaL..... 72
:::|||:|||||: |||
187 TATGCAGAACATTCTCATGACAAATGGTCAATGCACAAAGTTAGCAAAATGC 236

73 .....TyrlsGLUPheaRgVaLVaIGLUeuaSPProSeRalalysi 87
||| |||: ||| ::|||: |||
237 ATGGATTATTAGGAATA.....TATTCAAGACTCCTCCAAGA 274

87 leGU.....ValThrTYrTYraSPLYSanLSylsYLsgLUglU 100
||: ||| ||| ::|||: |||
275 TTC AAC CCTTGATGA AAC CATATA ACTATTATCTGA AAGAAAAAGAA 324

101 ThrLysSerPheProIleThrGLU 108
:::||| |||
325 ATT TAT CGCTG G CCAATCAAGAA 348

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-525-654A-2
seq_documentation_block:
; Sequence 2, Application US/08525654A
; Patent No. 5736356
; GENERAL INFORMATION:
; APPLICANT: SANO, KOHICHIRO
; APPLICANT: KUMAZAWA, YOSHIIYUKI
; APPLICANT: YASEUDA, HISASHI
; APPLICANT: SEGURO, KATSUYA
; APPLICANT: MOTOKI, MASAO
; TITLE OF INVENTION: TRANSGlutAMINase ORIGINATED FROM
; TITLE OF INVENTION: CRASSOSTREA GIGAS
; NUMBER OF SEQUENCES: 150
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/525,654A  
FILING DATE: 28-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6/8283  
FILING DATE: 28-JAN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7/3876  
FILING DATE: 13-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 10-760-0 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2310 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Crassostrea gigas  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..2310  
US-08-525-654A-2

alignment\_scores:  
Quality: 73.00 Length: 83  
Ratio: 1.404 Gaps: 4  
Percent Similarity: 62.651 Percent Identity: 25.301

alignment\_block:  
US-09-728-670-10 x US-08-525-654A-2 ..

Align seg 1/1 to: US-08-525-654A-2 from: 1 to: 2310

```
25 LeuMetValAsnValThrGlyValAspSerLysGlyAsnGluLeuLeu 41
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2062 CTCGTGCTCACCAGGCTGAATAAGCCGTGAGGGCTCGCTCTCCTAGC 2111
41 rProHisTyrValGluPhe.....ProIleLysProGly..... 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2112 TCCACAACCATAGATATAGCAGTCTCTATTAACTGTGAGCAGGTCA 2161
53 .....ThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAla 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2162 AGAAACAGTCGCTCTCGCCCGCCCAACCATACTAC.....TGGGGT 2205
68 LeuAspAlaThrAla...TyrLysGluPheArgValValGluLeuAsp 83
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2206 CGTGAAGTGAATGACGACCTTCACCTCCAAACAGATGTGCGACATTGAGAC 2255
83 oSerAlaLysIleGluValThrTyrTyrAspLysAsnLysLysLysGlu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2256 CAGCGCCGACATTAGGTGATCGACAGACAGAACAAAGATACGACAGCGAC 2304
```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-525-654A-4

seq\_documentation\_block:

Sequence 4, Application US/08525654A  
Patent No. 5736356  
GENERAL INFORMATION:  
APPLICANT: SANO, KOHICHIRO  
APPLICANT: KUMAZAWA, YOSHITUKI  
APPLICANT: YASEUDA, HISASHI  
APPLICANT: SEGURO, KATSUYA

APPLICANT: MOTOKI, MASAO  
TITLE OF INVENTION: TRANSGlutAMINASE ORIGINATED FROM  
TITLE OF INVENTION: CRASSOSTREA GIGAS  
NUMBER OF SEQUENCES: 150  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/525,654A  
FILING DATE: 28-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6/8283  
FILING DATE: 28-JAN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7/3876  
FILING DATE: 13-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 10-760-0 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2313 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Crassostrea gigas  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..2054  
US-08-525-654A-4

alignment\_scores:  
Quality: 73.00 Length: 83  
Ratio: 1.404 Gaps: 4  
Percent Similarity: 62.651 Percent Identity: 25.301

alignment\_block:  
US-09-728-670-10 x US-08-525-654A-4 ..

Align seg 1/1 to: US-08-525-654A-4 from: 1 to: 2313

```
25 LeuMetValAsnValThrGlyValAspSerLysGlyAsnGluLeuLeu 41
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2065 CTCGTGCTCACCAGGCTGAATAAGCCGTGAGGGCTCGGGTCTCCTAGC 2114
41 rProHisTyrValGluPhe.....ProIleLysProGly..... 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2115 TCCACAACCATAGATATAGCAGTCTCTATTAACTGTGAGCAGGTCA 2164
53 .....ThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAla 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2165 AGAAACAGTCGCTCTCGCCCGCCCAACCATACTAC.....TGGGGT 2208
68 LeuAspAlaThrAla...TyrLysGluPheArgValValGluLeuAsp 83
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```





```
; Patent No. 6287843
; GENERAL INFORMATION:
; APPLICANT: Baldwin, Donald A.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses
; FILE REFERENCE: 5718-44,
; CURRENT APPLICATION NUMBER: US/09/282,305
; CURRENT FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/080,563
; PRIOR FILING DATE: 1998-04-03
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1576
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (38)..(1336)
US-09-282-305-9
```

```
alignment_scores:
  Quality: 70.50      Length: 169
  Ratio: 0.892       Gaps: 11
  Percent Similarity: 46.746   Percent Identity: 23.669
```

```
alignment_block:
US-09-728-670-10 x US-09-282-305-9/rev ..
```

```
Align seg 1/1 to reverse of: US-09-282-305-9 from: 1 to: 1576
```

```
14 AspaLaserTyrPheGluProThrGlyProTyrLeuMetValAsnValTh 30
   ||| ||||| :||| ||||| ||||| :||| :||| :||| :||| :|||
988 GACCCAGTTTCAACAGCCCAACACCGTGTACATCTCTGTTGTTATC 939
30 rGlyValAspSer.....LysGlyAsn..GluLeuLeuSerProHisT 44
   : :||| :||| :||| :||| :||| :||| :||| :||| :|||
938 CACCACCTCCAGTTTACAGAGGGGAATATTGAATTCTTGACAAACTTT 889
44 YR..... 44
888 ACACATTCAGCATGGCTTCATATAGAGATTGAAGCCTTAACGATC 839
45 .....ValGluPheProIleLysProGlyThrThrLeu.... 55
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
838 CCTGCCCATGATCAGCCCAACATTTGAAGACATATGACACAGGAGAT 789
56 ..... 56
788 ATGTCTCAACAACCTTTGGCAATATTGTTTAAAGGCCGAGTAAAGCTA 739
59 ysIleGluTyrTyrVal..GluTrpAlaLeuAspAlaThr....Ala 72
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
738 GTGTCACTCTATCCCATCTTTAAGTGAATGTTGATGACATAATATTTC 689
73 TyrLysGluPheArgValValGluLeuAspProSerAlaLysIleGluVa 89
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
688 TTCCCTTTCTCTATATCTTAATATACACCTGTTCCAGAAAGAACAGGT 639
89 lThrTyrTyrAspLysAsnLysLysGluGluThrLysSer..PheP 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
638 CACCATACTT.....GTGGAACCTCACAGTCATTAAC 607
105 rolleThrGlu..LysGlyPheVal.....ValProAspLeu 116
   ||| :||| :||| :||| :||| :||| :||| :||| :|||
606 CTGTCACTGAATAAAGGCTTCTTCACTCCATCTTCATGATGAGACATC 557
117 SerGluHisIleLysAsnProGlyPheAsnLeuIleThrLysValValI 133
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
556 AATGTCAATATAAAGAACCTTGGC.....ATGTACT 525
```

```
133 eGluLys 135
   :||| |||
524 TGAGAG 518
```

```
seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-532-656-7
```

```
seq_documentation_block:
```

```
; Sequence 7, Application US/09532656
; Patent No. 6316608
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Mark A.
; APPLICANT: Ruvoilo, Michael
; APPLICANT: Arnold, Jr., Lyle J.
; TITLE OF INVENTION: COMBINED POLYNUCLEOTIDE SEQUENCES AS DISCRETE ASSAY ENDPOINTS
; FILE REFERENCE: IN-0017 US
; CURRENT APPLICATION NUMBER: US/09/532,656
; CURRENT FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 924
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6316608 YCF22
US-09-532-656-7
```

```
alignment_scores:
  Quality: 69.00      Length: 141
  Ratio: 1.030       Gaps: 7
  Percent Similarity: 47.518   Percent Identity: 27.660
```

```
alignment_block:
US-09-728-670-10 x US-09-532-656-7 ..
```

```
Align seg 1/1 to: US-09-532-656-7 from: 1 to: 924
```

```
5 AspLysGlyLysTyrLysLysGlyAspAspAlaSerTyrPheGluProTh 21
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
235 GACCTTTTCAAAATGAAAAAGTTGTTGACGCCGCTTATGCAAGAAAC 284
21 rGlyProTyrLeuMetValAsnValThrGlyValAspSerLysGlyAsn 38
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
285 GTTACTTTATGCAATGATGATGATTTGAAACTAATGTTGACAGACACT 334
38 lLeuLeuSerProHisTyrValGlu..... 46
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
335 CGGCTTAA...CCGATTGCCGTGAGGTCATCAACATATATTAACGACTGT 381
47 PheProIleLysProGlyThrThrLeuThrLysGluLysIleGlu..Ty 62
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
382 TTC.....AATTAAATTATATTAAGGTGAGAAATTT 413
62 rTyrValGluTrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValY 79
   ||| :||| :||| :||| :||| :||| :||| :||| :|||
414 GTATTGGAATGCGCAAGAAAGCAGACAACTAAAAAATCAAAAAAGAGTTG 463
79 alGluLeu.....AspProSerAlaLysIleGluValThrTyr 92
   ||| :||| :||| :||| :||| :||| :||| :||| :|||
464 TGCATATTGAGGCTGATGATGATGAGACGATGAAGATGTAGAGACTAT 513
93 AspLysAsnLysLysLysGluGluThrLysSerPheProIleThrGlu 109
   ||| :||| :||| :||| :||| :||| :||| :||| :|||
514 GATGACGAGAGATGAAGATGAAGAGATGACGAG..... 546
109 sGlyPheValValProAspLeuSerGluHisIleLysAsnPro..... 123
   ||| :||| :||| :||| :||| :||| :||| :||| :|||
547 .....GTTGTACCTGACTTAGTATGCTTAACCTGAACCAATTGTTG 589
124 .....GlyPheAsnLeuIle 128
   ||| :||| :||| :||| :||| :||| :||| :||| :|||
```

590 ACGAAGACGGTTTGAATTGGTT 612

seq\_name: /cgn2\_6/ptodata/1/lna/5B\_COMB.seq:US-08-795-475-4

seq\_documentation\_block:

Sequence 4, Application US/08795475

Patent No. 5965390

GENERAL INFORMATION:

APPLICANT: Bjvrck, Lars

APPLICANT: Sjbvrlng, Ulf

TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/795,475

FILING DATE: 11-FEB-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mcmasters, David D.

REGISTRATION NUMBER: 33,963

REFERENCE/DOCKET NUMBER: 100084.402D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1308 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Escherichia coli L392/pHDLG, DSM 7055

US-08-795-475-4

alignment\_scores:

Quality:	69.00	Length:	116
Ratio:	1.190	Gaps:	6
Percent Similarity:	50.000	Percent Identity:	31.034

alignment\_block:

US-09-728-670-10 x US-08-795-475-4 ..

Align seg 1/1 to: US-08-795-475-4 from: 1 to: 1308

5 AsplysGlyLysTyrLysGlyAspAspAlaSerTyrPheGluProTh 21

823 GAAATGTAATAT.....ACAGCAGACTTAGAAGATCG 857

21 rGlyProTyrLeuMetValAsnValThrGly.....ValAspSerLysG 36

858 TGGATACACTATTAATATTAAGATTGCAAGTAAGAAAGTTGACGAAAAAC 907

36 lYAsnGluLeuLeuSerProHisTyrValGluPheProIleLysProGly 52

908 CAGAGAACCACATGACACTTACAAATTA.....ATCCTAATGCT 948

53 ThrThrLeuThrLysGluLysIleGluTyrTyrValGluTyrPalaeuAs 69

949 AAACATTTGAAGCGCAAACTACTGAAGCTGTGAT.....GC 989

69 pAlaThrAlaTyrLysGluPheArgValAlaGluLeuAspProSerAla 86

990 TGCTACTGCAGAAAAAGCTTCAACAATACGCTAACGACACGCTGTG 1039

86 ySileGluValThrTyrTyrAspLysAsnLysLysGluGluThrLys 102

1040 ACGGTGAATGACTTACGACGAT.....GCGACTAAG 1071

103 SerPheProIleThrGluLysGlyPheValValProAspLeuSerGlu 118

1072 ACCTTACAGTTACTGAAAAACAGAAAGTATC..GATGCGTCTGAA 1116

seq\_name: /cgn2\_6/ptodata/1/lna/6B\_COMB.seq:US-09-064-693A-22

seq\_documentation\_block:

Sequence 22, Application US/09064693A

Patent No. 6210937

GENERAL INFORMATION:

APPLICANT: Ward, Thomas E.

TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY

TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION

TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: W. Gary Goodson

ADDRESSEE: INEEL--Lockheed Martin Idaho

ADDRESSEE: Technologies Co.

STREET: P.O. Box 1625

CITY: Idaho Falls

STATE: Idaho

COUNTRY: USA

ZIP: 83415-3810

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage

COMPUTER: Toshiba Satellite Pro T2150CDS

OPERATING SYSTEM: Windows95

SOFTWARE: Word Perfect 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/064,693A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: W. Gary Goodson

REGISTRATION NUMBER: 22,387

REFERENCE/DOCKET NUMBER: LIT-PI-296

TELECOMMUNICATION INFORMATION:

TELEPHONE: (208)526-9469

TELEFAX: (208)526-8339

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 1348 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-064-693A-22

alignment\_scores:

Quality:	69.00	Length:	106
Ratio: <td>1.302<td>Gaps:<td>4</td></td></td>	1.302 <td>Gaps:<td>4</td></td>	Gaps: <td>4</td>	4
Percent Similarity: <td>50.000<td>Percent Identity:<td>24.528</td></td></td>	50.000 <td>Percent Identity:<td>24.528</td></td>	Percent Identity: <td>24.528</td>	24.528

alignment\_block:

US-09-728-670-10 x US-09-064-693A-22 ..

Align seg 1/1 to: US-09-064-693A-22 from: 1 to: 1348

11 LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVa 27

```
620 AAAGCGATCTGTTTCT.....AAACCTATATCGACAT 654
27 IAsnValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHist 44
   :::::::::: :::: ||| ||::: :::: ||||
655 CACACTCAATCTGATGAAGACGTTGGTGTGAATGAAATCAGACACT 704
44 yRValGluPheProIleLys.....ProGlyThr 53
   || ::||| ::||| ||||| ||||| |||||
705 ATCAACAATTTGTCGTAAGAGCGGCGACGCTTATCAGTCTCCGGGTACT 754
54 ThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAspAl 70
   ||| ||||| |||||
755 .....TATTGGTCGAAGC.....GATGC 774
70 aThrAlaTyrLysGluPheArgValValGluLeuAspProSerAlaLysI 87
   |::: :::: ||| ::::: ::::: :::::
775 ATCTTCGGCTCTTACTTCTTCTGGCAGCAGCAGCAATCAAGCGGCGACTG 824
87 IeGluValThrTyrTyrAspLysAsnLysLysLysGluGluThrLysSer 103
   ::::||| ||| ::||| :::: ::::
825 TAAAGTGACCGGTATTGACGTAACAGTATGACGCGGTGATTCGCTTT 874
104 PheProIleThrGluLys 109
   :: |||||
875 GCTGATGTGCTGAAAAA 892

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-064-693A-26

seq_documentation_block:
; Sequence 26, Application US/09064693A
; Patent No. 6210937
;
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Gary Goodson
; ADDRESSEE: INEEL--Lockheed Martin Idaho
; ADDRESSEE: Technologies Co.
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: Toshiba Satellite Pro T2150CDS
; OPERATING SYSTEM: Windows95
; SOFTWARE: Word Perfect 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,693A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: W. Gary Goodson
; REGISTRATION NUMBER: 22,387
; REFERENCE/DOCKET NUMBER: LIT-PI-296
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (208)526-9469
; TELEFAX: (208)526-8339
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4530 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-09-064-693A-26
```

```
alignment_scores:
      Quality: 69.00      Length: 106
      Ratio: 1.302      Gaps: 4
      Percent Similarity: 50.000      Percent Identity: 24.528

alignment_block:
US-09-728-670-10 x US-09-064-693A-26 ..
Align seg 1/1 to: US-09-064-693A-26 from: 1 to: 4530

11 LysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetVa 27
   ||||| ||| ::||| ||||| ::::
2502 AAAGCGATCTGTTTCT.....AAACCTATATCGACAT 2536
27 IAsnValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHist 44
   :::::::::: :::: ||| ||::: :::: ||||
2537 CACACTCAATCTGATGAAGACGTTGGTGTGAATGAAATCAGACACT 2586
44 yRValGluPheProIleLys.....ProGlyThr 53
   || ::||| ::||| ||||| ||||| |||||
2587 ATCAACAATTTGTCGTAAGAGCGGCGACGCTTATCAGTCTCCGGGTACT 2636
54 ThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAspAl 70
   ||| ||||| |||||
2637 .....TATTGGTCGAAGC.....GATGC 2656
70 aThrAlaTyrLysGluPheArgValValGluLeuAspProSerAlaLysI 87
   |::: :::: ||| ::::: ::::: :::::
2657 ATCTTCGGCTCTTACTTCTTCTGGCAGCAGCAGCAATCAAGCGGCGACTG 2706
87 IeGluValThrTyrTyrAspLysAsnLysLysLysGluGluThrLysSer 103
   ::::||| ||| ::||| :::: ::::
2707 TAAAGTGACCGGTATTGACGTAACAGTATGACGCGGTGATTCGCTTT 2756
104 PheProIleThrGluLys 109
   :: |||||
2757 GCTGATGTGCTGAAAAA 2774

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-08-960-780-26

seq_documentation_block:
; Sequence 26, Application US/08960780
; Patent No. 6204435
;
; GENERAL INFORMATION:
; APPLICANT: Feltelson, Jerald S.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schneits, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: Stamp, Lisa
; TITLE OF INVENTION: No. 6204435e1 Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,780
; FILING DATE: 30-OCT-1997
; CLASSIFICATION: 435
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA-708
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1278 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 68F
; US-08-960-780-26
```

```
alignment_scores:
      Quality: 68.50      Length: 153
      Ratio: 0.867      Gaps: 8
      Percent Similarity: 51.634      Percent Identity: 20.261
```

```
alignment_block:
US-09-728-670-10 x US-08-960-780-26 ..
```

```
Align seg 1/1 to: US-08-960-780-26 from: 1 to: 1278
```

```
10 LysLysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMe 26
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 AAAGAAGCAGCGGCGATTTCATTACCTTATCAGATGATGACATACGAT 216
26 tValAsnValThrGly.....ValAspSerLysGlyAsnGluLeuLeus 41
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
217 GATGAAATCGATGCGAAAGTTATTCTTAATTAAGCGAAAGAAACAAG 266
41 erProHisTyrValGluPheProIleLysProGlyThrThrLeuThrLys 57
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
267 TTGTCCAT.....TTAGAAAAGGA...CAGTTCGTTCT 298
58 GluLysIleGluTyrTyrValGlu.....TyrPalaLeuAspAlaThrAl 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
299 ATCAAAATAGATATCAAGCTGATGACCATTTAATGCGGATAGTCAAC 348
72 aTyrLysGluPheArgValValGluLeuAspProSerAlaLysIleGluV 89
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
349 CTTTAAAAATTTGAACCTCTTTAAAGTAGATTAAGCAACAGTCCACG 398
89 aLThrTyrTyrAspLys.....AsnLysLysLysGlu 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
399 AAATTCACCTAGATGATTAAGAAACCTGAAATTTAATAAAAAGAAACA 448
100 Glu.....ThrLysSerPheProIleThrGluLys.. 109
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
449 CAAGAATTTCTAACAAAGCAACAAAACAACTTATTACTCAAAAAGT 498
110 .....GlyPheValValProA 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
499 GAAGAGTACTAGGATGAAGACACGATACAGATTCATATCCAG 548
115 splLeuSerGluHisIleLysAsnProGlyPheAsnLeuIleThrLysVal 131
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
549 ACATTTGGGA.....GAAAATGGGTATACCATCCAAATTAAGATT 589
132 ValIleGlu 134
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
590 GCCGTCAAA 598
```

```
seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-073-898-26
```

```
seq_documentation_block:
; Sequence 26, Application US/09073898
; Patent No. 6242669
; GENERAL INFORMATION:
; APPLICANT: Feltelson, Jerald S.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeltz, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: stamp, Lisa
; APPLICANT: Morrill, George
; APPLICANT: Finstad-Lee, Stacey
; TITLE OF INVENTION: No. 6242669el Pesticidal Toxins and Nucleotide
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,898
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/960,780
; FILING DATE: 30-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: MA-708C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1278 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 68F
; US-09-073-898-26

alignment_scores:
      Quality: 68.50      Length: 153
      Ratio: 0.867      Gaps: 8
      Percent Similarity: 51.634      Percent Identity: 20.261

alignment_block:
US-09-728-670-10 x US-09-073-898-26 ..

Align seg 1/1 to: US-09-073-898-26 from: 1 to: 1278

10 LysLysGlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMe 26
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 AAAGAAGCAGCGGCGATTTCATTACCTTATCAGATGATGACATACGAT 216
```



```
26 tValAsnValThrGly.....ValAspSerLysGlyAsnGluLeuLeus 41
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
217 GATGAAATCGATGCGAAGTATTCTTAATTAAGGGAAGAAAGAAACAG 266
41 erProHisTyrValGluPheProIleLysProGlyThrThrLeuThrLys 57
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
267 TTGTCAT.....TTAGAAAAGGA...CAGTTCGTTCT 298
58 GluLysIleGluTyrTyrValGlu.....TrpAlaLeuAspAlaThrAl 72
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
299 ATCAAAATAGAAATCAAGCTGATGAACCATTTAATCGCGATAGTCAAC 348
72 aTyrLysGluPheArgValValGluLeuAspProSerAlaLysIleGluV 89
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
349 CTTTAAAAATTTGAACCTCTTTAAAGTAGATAGTAAACACAGTCCAGC 398
89 aThrTyrTyrAspLys.....AsnLysLysLysGlu 99
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
399 AAATTCACATGATGATTAAGAAACCCCTGAATTTAATAAAAGAAACA 448
100 Glu.....ThrLysSerPheProIleThrGluLys.. 109
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
449 CAAGAATTCTACAAAGCAACAAACAAACCTTATTACTCAAAAGT 498
110 .....GlyPheValPro 115
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
499 GAAGAGTACTAGGATGAGACACGAGATACAGATGAGATCTATTCCAG 548
115 spleuSerGluHisIleLysAsnProGlyPheAsnLeuIleThrLysVal 131
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
549 ACATTTGGGA.....GAAATGGGTATACCATCCAAATAAGATT 589
132 ValIleGlu 134
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
590 GCCGTCAA 598

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-257-073-10

seq_documentation_block:
; Sequence 10, Application US/08257073
; Patent No. 576597
; GENERAL INFORMATION:
; APPLICANT: Paolletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
```

```
ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-257-073-10

alignment_scores:
Quality: 68.50 Length: 103
Ratio: 1.245 Gaps: 6
Percent Similarity: 53.398 Percent Identity: 30.097

alignment_block:
US-09-728-670-10 x US-08-257-073-10 ..

Align seg 1/1 to: US-08-257-073-10 from: 1 to: 5181

28 AsnValThrGlyValAspSerLysGlyAsnGluLeuSerProHisTyr 44
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1021 AACATTGAAGAAATTACTTGATAGATTAATGAATTAATAATCCACC 1070
44 rValGluPheProIleLysProGlyThrThrLeuThrLys..GluLysI 60
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1071 GCCCAATTCGGAATACACCAATATCTCTCTTGATAGACAAAAAAA 1120
60 leGluTyrTyrValGluTTrpAlaLeuAspAlaThrAlaTyrLysGluPhe 76
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1121 TCGAGAACACGAAAGAAAAATAAAGAAATTCGCAAACTAATAATT 1170
77 ArgValValGluLeu.....AspProSerAlaLysIleGluValThrTy 91
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1171 AACATTGATGATTATTACTGATCA.....CTTGAATTAAGATA 1211
91 rTyr.....AspLysAsnLysLys.....LysG 99
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1212 TTATTTAAGAGAAAAATAAAGTTGATGTACACCTTAATCAACAG 1261
99 lGluThrLysSerPheProIleThrGlu.....LysGlyPhe 111
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1262 ATCTACGAATCTGTCAATATACCAAAAGTTCCTTATCCAAATGGTATT 1311
112 ValValPro 114
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1312 GTATATCT 1320

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-198-956-1

seq_documentation_block:
; Sequence 1, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schmitt, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377,200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
```





1194 TGTGTGTCATGCCAGCGCCCATACAGTGACGCCATTTCGAATGGA 1243  
82 P.....ProserAlaLysIleGluValThrTyrTyrAspLysAsnLysL 97  
1244 CCGCAACCCATCCGGTCTGCCGACTCCCGCTAAACGACACTCCAAGA 1293  
97 yslYsgLgluLthrLysSerPheProIleThrGluLysGlyPheValVal 113  
1294 TCAAGTCGACTTTCATCAATCTCCGTCATGACGCCGAAATGGTCATC 1343  
114 ProAspLeuSerGluHisIleLysAsnProGlyPhe 125  
1344 CCATCTATTGGCGTTCACATCGTATTGCCCTCTTTT 1379

seq\_name: /cgn2\_6/ptodata/1/lna/6B\_COMB.seq:US-09-297-928-3

seq\_documentation\_block:

; Sequence 3, Application US/09297928  
; Patent No. 6358716  
; GENERAL INFORMATION:  
; APPLICANT: BULTHUIS, BEN A.  
; GATENBY, ANTHONY A.  
; HAYNIE, SHARON L.  
; HSU, AMY K.  
; LAREAU, RICHARD D.  
; TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF  
; GLYCEROL BY RECOMBINANT  
; ORGANISMS  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENENCOR INTERNATIONAL, INC.  
; STREET: 4 CAMBRIDGE PLACE  
; 1870 SOUTH WINTON ROAD  
; CITY: ROCHESTER  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 14618  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD VERSION 7.0A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/297, 928  
; FILING DATE: 11-May-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/03602  
; FILING DATE: NOVEMBER 13, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FLOYD, LINDA AXAMETHY  
; REGISTRATION NUMBER: 33,692  
; REFERENCE/DOCKET NUMBER: CR-9981-P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-892-8112  
; TELEFAX: 302-773-0164  
; TELEX: 6717325  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3178 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-297-928-3

alignment\_scores:  
Quality: 67.00 Length: 112  
Ratio: 1.063 Gaps: 4  
Percent Similarity: 56.250 Percent Identity: 23.214

alignment\_block:

US-09-728-670-10 x US-09-297-928-3 ..

Align seg 1/1 to: US-09-297-928-3 from: 1 to: 3178

29 ValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHisTyrVa 45  
1053 ATCAGCGGTGTGAGAACGCGCTACCGTCTTGATC.....TATGT 1093  
45 IgluPhePro.....IleLysProGlyThr..... 54  
1094 CGAGGTACAAATAATTGATCAAGACCCCAACTTCTGTTAGCTTATCGGTG 1143  
55 .....LeuThrLysGluLysIleGluTyrTyrValGlu 65  
1144 CCGAGGCCCGGACGTTGAGACTTAATGAGCTTGTACGAATCAACGCTAA 1193  
66 TrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValGluLeuAs 82  
1194 TGTGTGTCATGCCAGCGGCCCATACAGTGACGCCCATTTTGCAAATGGA 1243  
82 P.....ProserAlaLysIleGluValThrTyrTyrAspLysAsnLysL 97  
1244 CCGCAACCCATCCGCTCTGCCGACTCCCGCTAAACGACACTCCAAGA 1293  
97 yslYsgLgluLthrLysSerPheProIleThrGluLysGlyPheValVal 113  
1294 TCAAGTCGACTTTCATCAATCTCCGTCATGACGCCGAAATGGTCATC 1343  
114 ProAspLeuSerGluHisIleLysAsnProGlyPhe 125  
1344 CCATCTATTGGCGTTCACATCGTATTGCCCTCTTTT 1379

seq\_name: /cgn2\_6/ptodata/1/lna/6B\_COMB.seq:US-09-371-913A-10

seq\_documentation\_block:

; Sequence 10, Application US/09371913A  
; Patent No. 6297369  
; GENERAL INFORMATION:  
; APPLICANT: Schnepf, H. Ernest  
; APPLICANT: Narva, Kenneth E.  
; APPLICANT: Stockhoff, Brian A.  
; APPLICANT: Finstad Lee, Stacey  
; APPLICANT: Walz, Mikki  
; APPLICANT: Sturgis, Blake  
; TITLE OF INVENTION: Pesticidal Toxins and Genes from Bacillus laterosporus  
; FILE REFERENCE: MA-719XC2 US  
; CURRENT APPLICATION NUMBER: US/09/371, 913A  
; CURRENT FILING DATE: 1999-08-10  
; PRIOR APPLICATION NUMBER: 60/095, 955  
; PRIOR FILING DATE: 1998-08-10  
; PRIOR APPLICATION NUMBER: 60/138, 251  
; PRIOR FILING DATE: 1999-06-08  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 1041  
; TYPE: DNA  
; ORGANISM: Bacillus laterosporus  
US-09-371-913A-10

alignment\_scores:  
Quality: 66.50 Length: 176  
Ratio: 0.899 Gaps: 9  
Percent Similarity: 42.045 Percent Identity: 21.023

alignment\_block:

US-09-728-670-10 x US-09-371-913A-10 ..

Align seg 1/1 to: US-09-371-913A-10 from: 1 to: 1041





```
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Antiviral Rictin-Like Proteins
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/147,208
; FILING DATE: 02-MAR-1999
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Rudolph, John R.
; REGISTRATION NUMBER: 38,003
; REFERENCE/DOCKET NUMBER: 7841-76
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9639 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-147-208-26

alignment_scores:
      Quality: 66.00      Length: 138
      Ratio: 0.930      Gaps: 11
      Percent Similarity: 51.449      Percent Identity: 26.812

alignment_block:
US-09-728-670-10 x US-09-147-208-26/rev ..

Align seg 1/1 to reverse of: US-09-147-208-26 from: 1 to: 9639

      8 LysTYrLysLysGlyAspAspAlaSerTYrPheGluProThr..... 21
      |||||
      1438 AAATACAAA.....TCGTTATTGGAGCAGACCATCCAAAA 1404

      22 ....GlyProTYrLeuMetValAsnValThrGlyValAspSerLysGlyA 37
      |||||
      1403 TGGGGGCCCT.....ACCTCTAACAGCGTCGTGAAAAAGGGCG 1366

      37 sngluleu..LeuSerProHisTYrValGluPheProIleLysProGly 52
      ::||| |||::||| ||| ::||| |||
      1365 ACCCGCTGATTGGACCGCATACAGTG..CTCATTTACTAATCCGCG 1319

      53 Thr.....ThrLeuThrLysGluLysIleGluTYrTYrVal..... 64
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
      1318 GTGATTTCAGCTGATTATGCAAGTCCAAATTCCTTACGCCCATAGAAATTACA 1269

      65 .GluTrpAlaLeuAspAlaThrAlaTYrLysGluPheArgValGluL 81
      ||||| |||::|||::|||::|||::|||::|||::|||::|||
      1268 AGAATGGCTTTAGAGAGTAATTCCTCAAGTCCTGCACCCGCAAGT 1219

      81 euaspProSerAlaLysIleGluValThrTYrTYrAspLysAsnLysLys 97
      |||||::|||::|||::|||::|||::|||::|||::|||
      1218 ACGATCCCGCATCAA.....CAACGG 1196

      98 LysGluGluThrLysSerPheProIleThrGluLysGlyPheValValPr 114
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
```

```
1195 GAGGAGGAGAGCAACAGTTC...GTGACTAAG.....CTGATTGC 1158
1114 cAspleuSerGluHisIleLysAsnProGlyPheAsnLeuIleThrLysV 131
      ::|||::|||::|||::|||::|||::|||::|||::|||
1157 GACATTCAACCGAGCACACA.....AACCGCTGCAAGCGG 1123
      ||||| ::|||
1122 TGTGGCGCAAAA 1109

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:US-08-978-741-5

seq_documentation_block:
; Sequence 5, Application US/08978741
; Patent No. 610076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 610076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11284 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; US-08-978-741-5

alignment_scores:
      Quality: 66.00      Length: 138
      Ratio: 0.930      Gaps: 11
      Percent Similarity: 51.449      Percent Identity: 26.812

alignment_block:
US-09-728-670-10 x US-08-978-741-5/rev ..

Align seg 1/1 to reverse of: US-08-978-741-5 from: 1 to: 11284

      8 LysTYrLysLysGlyAspAspAlaSerTYrPheGluProThr..... 21
      |||||
      1438 AAATACAAA.....TCGTTATTGGAGCAGACCATCCAAAA 1404

      22 ....GlyProTYrLeuMetValAsnValThrGlyValAspSerLysGlyA 37
      |||||
      1403 TGGGGGCCCT.....ACCTCTAACAGCGTCGTGAAAAAGGGCG 1366

      37 sngluleu..LeuSerProHisTYrValGluPheProIleLysProGly 52
```

```

      1365 ACCCGCTGATTTGCAGCCGCATACAGTGT...CTCATTACTAAATCCGGC 1319
      53 Thr.....ThrlleuThrlsGluLysIleGluTyrTyrVal..... 64
      1318 GTGATTCAGCTGATTAAGAAGTCCAAATGCTTACGCCATAGAAATACA 1269
      65 .GluTrpAlaLeuAspAlaThrAlaTyrTrpGluPheArgValGlu 81
      1268 AGAATGGCTTTAGAGAAGTAATTCCTCAAGTGTCTGCACCGGCAAGT 1219
      81 euAspProSerAlaLysIleGluValThrTyrTrpAspLysAsnLys 97
      1218 ACGATCCCGCGATCAAA.....CAACGG 1196
      98 LysGluGluThrLysSerPheProIleThrGluLysGlyPheValValPr 114
      1195 GAGGAGGAGAGCAACAGTTG...GTGACTAAG.....CTGATTGC 1158
      114 oAspLeuSerGluHisIleLysAsnProGlyPheAsnLeuIleThrLysV 131
      1157 GACATTTCACCGAGCACACA.....AACGCGCTGCAAGCGG 1123
      131 alValIleGluLys 135
      1122 TGGTGGCGCAAAA 1109

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-852-481-1

seq_documentation_block:
; Sequence 1, Application US/08852481
; Patent No. 5928931
; GENERAL INFORMATION:
; APPLICANT: Grun Ph.D., Felix
; APPLICANT: Buck Ph.D., Jochen
; APPLICANT: Hammerling Ph.D., Ulrich
; TITLE OF INVENTION: ISOLATION, PURIFICATION AND CLONING OF
; TITLE OF INVENTION: RETINOL DEHYDRATASE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,481
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/017,178
; FILING DATE: 09-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/771
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1564 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-852-481-1
```

```

alignment_scores:
      Quality: 65.50      Length: 125
      Ratio: 1.074      Gaps: 7
      Percent Similarity: 48.800      Percent Identity: 23.200

alignment_block:
US-09-728-670-10 x US-08-852-481-1 ..
Align seg 1/1 to: US-08-852-481-1 from: 1 to: 1564

23 ProTyrLeu.....MetValAsnValThrGlyValAspSerLysGlyAs 37
   |||||
227 CCTTACTTGAAGATGCGCGCGAATATCTACACATGCCCTTAAGACCTAC 276
37 nGluLeuLeuSerProHisTyrValGluPheProIleLysProGlyThrT 54
   |||||
277 AGACGTGTTGCTTGCAGTTATCAA.....CGATCAGGAACGA 314
54 hrLeuThrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAsp... 69
   |||||
315 CAATGACTCAAGAACTAGT.....TGGCTAATTGAAAAC 349
70 .....AlaThrAlaTyrLysGluPheArgValGlu 80
350 GACTTGAAATTCGAAGCTGCAAAACATACATGTCCCTCCGCTACATTTA 399
80 uLeuAspProSerAlaLysIleGluValThrTyrTrpAspLysAsnLysL 97
   |||||
400 TCTTGAC.....GGCTTCATGATCTACGACCCGGAAGAC 434
97 yLysGluGluThrLysSerPheProIleThrGluLys..... 109
   |||||
435 AAGAAGATATAACGACATATACCAAAATCCAGAAACCTTGATATGGA 484
110 .....GlyPheValValProAspLeuSerGluHisIleLysAsnPr 123
485 AGGTATTAGGA.....TTGCTAGAATACTTTAGTGTCC 519
123 oGlyPheAsnLeuIleThrLysVal 131
520 AGGAGCTCATTTGCTGCTGCAGTGTG 544

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-232-079-1

seq_documentation_block:
; Sequence 1, Application US/08232079
; Patent No. 5500361
; GENERAL INFORMATION:
; APPLICANT: Kinney, Anthony J.
; TITLE OF INVENTION: Nucleotide Sequences of
; TITLE OF INVENTION: Soybean beta-ketoacyl-ACP
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. du Pont de Nemours and
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.50 Diskette
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,079
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: 07/791,921  
FILING DATE: 15 NOVEMBER 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Slegell, Barbara C.  
REGISTRATION NUMBER: 30,684  
REFERENCE/DOCKET NUMBER: BB-1035-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (302) 992-4931  
TELEFAX: (302) 773-0164  
TELEX: 835420  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2675 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Glycine max  
STRAIN: Cultivar Wye  
CELL TYPE: Cotyledon  
IMMEDIATE SOURCE:  
LIBRARY: cDNA to mRNA  
CLONE: pC161  
US-08-232-079-1

alignment\_scores:  
Quality: 65.50 Length: 191  
Ratio: 0.851 Gaps: 5  
Percent Similarity: 40.314 Percent Identity: 19.372

alignment\_block:  
US-09-728-670-10 x US-08-232-079-1 ..

Align seg 1/1 to: US-08-232-079-1 from: 1 to: 2675

5 AsplysGlyLySTyRlySlySglyAsp.....AspAl 15  
|||::: ::|||::|||::|||  
1436 GACCGTGATCAGTATTAACAAGCTGATTTCCAGAAGATATACGATGC 1485  
15 aSerTyRphe.....GluProThrGlyProTyRL 25  
|:::||||  
1486 TAAATTTTTCATCAAGTCTTACAGTGAGGATGATATCCATTAAGATA 1535  
25 euMetValAsnValThrGlyValAspSerLysGlyAsnGluLeuLeuSer 41  
::: ||||| ::: ::|||::|||::|||  
1536 TTAAGTATAACGCTGCGCCAGTACCCCAAAATGGCAACAAGAGCTTGAT 1585  
42 ProHisTyValGluPheProIleLysProGlyThrThrLeuThr.... 56  
||| ||| ||| ||| ||| |||  
1586 GCTGCATATCATGAGGCACACAGAAACCTGGTGGCTGCCCTGTTTCT 1635  
56 ..... 56  
1636 CTTTTCCTCAGTTAATACCAAGTGGGCAATTGTGGGGCTGCTGAATGA 1685  
57 .....LysGluLysIleGluTyRThrValGluTTPAla 67  
::: ||||| ::|||::|||::|||  
1686 TTGCTCCTGTTGATTTTAAACAAGAGTGTGAGTATTGGCAGCAAGCAAG 1735  
68 LeuAspAlaThrAlaTyRlySgluPheArgValValGluLeuAspProse 84  
::: ||||| ::|||::|||::|||  
1736 TCGAATGGCTGCTTCTCTTAAGTGCGACGTTGTTAAGGATGTTCTTAA 1785  
84 rAlaLysIleGluValThrTyRlyAspLysAsnLysLysGluGluT 101  
: ||||| ::|||::|||::|||  
1786 CAATTTGTGAGGCACATTAACCTGATTAACAATGAGTACAACCTGTCA 1835  
101 hrLysSerPheProIleThrGluLysGlyPheValValProAspLeuSer 117  
|||::|||

1836 CTACAGT.....AGGATACACAA 1855  
118 GluHisIleLysAsnProGlyPheAsnLeuIle..... 128  
||| ::|||::|||::|||  
1856 GAGGTAATGTGAGGCCCTGGCCCTGAATTAATCAAAATTTCAGGAATA 1905  
129 .....T 129  
1906 TACCACAAGACATGCATTCGTGATGATTTTGGCTTCTATGAGGCCGCTC 1955  
129 hrLysValValIleGluLysLys 136  
|||:::|||||  
1956 AGAAGACTATTTTGGAGAGAAA 1978



---

4

.

OM of: US-09-728-670-10 to: EST:\* out\_format : pfs

Date: Sep 1, 2002 3:56 AM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

#### Command line parameters:

```
-MODEL=framet_p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09728670/runat_30082002_084912_5292/app_query.fasta.1.196
-DB=EST -DEMT=fastap -SUFFIX=first -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blsum62 -TRANS=human40.cdl
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=50 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09728670_@CGN1_1_2488
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1
```

#### Search information block:

```
Query: US-09-728-670-10
Query length: 136
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1752.300000
```

#### score\_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_est2:BI207534	+	85.00	7.51	655	BI207534 EST525574 CTOS Lycoper
gb_est1:AW485868	-	83.00	163.41	393	AW485868 68470 MARC 1Pig Sus sc
gb_gss:AZ052631	-	82.00	159.39	500	AZ052631 RPT-23-413D16.TJ RPT
gb_gss:BH244865	+	82.00	156.10	747	BH244865 AUIJA33TF AUIJ Arabid
gb_est2:BG584867	+	81.50	154.59	793	BG584867 EST486628 MHAM Medica
gb_est2:BI203214	+	80.00	153.50	622	BI203214 EST521254 CTOS Lycoper
gb_est2:BI769234	+	80.00	148.83	1100	BI769234 603060063F1 NIH_MGC_1
gb_est2:BG947905	+	79.00	151.71	602	BG947905 IPL_6_G12.B1.A002 Imma
gb_est2:BE98438	+	78.50	150.92	585	BE98438 EST430161 GVSN Medica
gb_est2:BI953150	+	78.00	151.50	481	BI953150 HVSME0011F14f Hordeu
gb_est2:BG581416	+	78.00	148.33	708	BG581416 EST483149 GVN Medica
gb_est2:BE455199	+	78.00	148.30	711	BE455199 HVSME0096J04f Hordeu
gb_est2:BF260862	+	78.00	148.05	733	BF260862 HVSMEf0023C06f Hordeu
gb_est2:BG583222	+	78.00	145.04	734	BG583222 EST484972 GVN Medica
gb_gss:CNS072XD	+	78.00	145.46	1006	AL426791 clone BA0AB015B05 of
gb_gss:CNS073BU	+	78.00	145.27	1030	AL427312 clone BA0AB018D09 of
gb_gss:BH483792	+	77.50	146.16	815	BH483792 BOGWK46TR BOGW Brassic
gb_est2:BE517562	-	76.50	147.92	54.17	BE517562 WHE0628_F09.L182A whea
gb_gss:AO501650	-	76.00	147.99	53.69	AO501650 V17E9 mTn-3xHA/lacZ in
gb_est1:AA840229	+	76.00	147.88	453	AA840229 vw89a06.r1 Stratagene
gb_gss:AZ672674	+	76.00	142.41	885	AZ672674 ENTTLT05TR Entamoeba h
gb_gss:BH158255	-	75.50	142.36	890	BH158255 ENTRTL14TF Entamoeba h
gb_est2:BE536273	-	75.50	141.98	822	BE536273 601062663F1 NIH_MGC_10
gb_est1:AU052970	+	75.00	145.36	480	AM036063 EST278077 tomato seed,
gb_est2:BM474008	+	75.00	144.20	553	AU052970 AU052970 Dictyostelium
gb_est2:AA549478	+	75.00	138.71	1082	BM474008 AGENCOURT_6489861 NIH
gb_est1:AA549478	-	74.50	146.29	378	AA549478 VK86f12.s1 Knowles Sol
gb_est2:BI204379	-	74.50	143.65	522	BI204379 EST522419 CTOS Lycoper
gb_est2:BI204935	-	74.50	143.65	522	BI204935 EST522975 CTOS Lycoper
gb_est2:BI206210	-	74.50	143.65	522	BI206210 EST524250 CTOS Lycoper
gb_est2:BI206708	-	74.50	143.65	522	BI206708 EST524748 CTOS Lycoper
gb_est2:BI207927	-	74.50	143.65	522	BI207927 EST525967 CTOS Lycoper
gb_est2:BI211112	-	74.50	143.65	522	BI211112 EST529152 CTOS Lycoper
gb_est2:BI955134	+	74.50	141.50	679	BI955134 HVSME0021119f Hordeu
gb_est2:BF570018	-	74.50	138.39	993	BF570018 602186015T1 NIH_MGC_45
gb_est1:AW036324	+	74.00	143.26	483	AW036324 EST278449 tomato seed,
gb_est1:AW036351	+	74.00	143.17	488	AW036351 EST278476 tomato seed,
gb_est2:BI391530	+	74.00	141.31	613	BI391530 pgp1n.pk004.119 Normal
gb_est2:BI065721	+	74.00	140.78	654	BI065721 pgf1n.pk005.k14 normal
gb_est2:BI929102	+	74.00	140.01	718	BI929102 EST548991 tomato flowe

gb_gss:AZ671442	-	73.50	137.80	198.44	830	1	AZ671442 ENTLY12TR Entamoeba
gb_est1:AI164607	+	73.00	143.72	92.86	355	1	AI164607 A065P71U Hybrid asp
gb_est2:T19018	+	73.00	142.71	105.79	402	1	T19018 f11010t Testis 1 Homo
gb_est2:BF158099	+	73.00	141.66	121.03	457	1	BF158099 f128e10.y1 Sugano K
gb_est1:AL389514	+	73.00	140.49	140.55	527	1	AL389514 MtBC55D02R1 MtBC Me

seq\_name: gb\_est2:BI207534

#### seq\_documentation\_block:

LOCUS BI207534 655 bp mRNA linear EST 11-JUL-2001  
DEFINITION EST525574 CTOS Lycopersicon esculentum cDNA clone cTOS14114 5' end,  
mRNA sequence.  
ACCESSION BI207534  
VERSION BI207534.1 GI:14685258  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 655)

#### REFERENCE

##### AUTHORS

van der Hoeven, R., Sun, H., Bezzerides, J., Cho, J., Utterback, R.,  
Ronning, C. and Tanksley, S.

##### JOURNAL

Generation of ESTs from Tomato Suspension Cultures  
Unpublished (2001)

##### COMMENT

Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>.

#### FEATURES

##### source

1.655  
/organism="Lycopersicon esculentum"  
/cultivar="TA96, E6203"  
/db\_xref="taxon:4081"  
/clone="cTOS14114"  
/clone\_11b="CTOS"  
/tissue\_type="suspension cultures"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Suspension cultures of L.esculentum E6203 were grown  
in Murashige and Skoog based medium, supplemented with 15%  
coconutmilk (filter sterilized and added after  
autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).  
Fresh medium was added every 7 days, and cultures were  
grown at 25 C, with 12hrs of light and continuous  
shaking."

BASE COUNT 200 a 147 c 132 g 176 t  
ORIGIN

#### alignment\_scores:

Quality:	Ratio:	Length:
Ratio: 1.232	6	136
Percent Similarity: 50.735	Percent Identity: 26.471	

#### alignment\_block:

US-09-728-670-10 x BI207534 ..

Align seg 1/1 to: BI207534 from: 1 to: 655

20 ProThrglyProTyrluMetValAsnValThrglyValAspserlysc1 36  
|||||

270 CCACAGGGAATTGCCCGCTCAAGACACACCCCTATGAT..... 311  
|||

36 yAsnGluLeuSerProHstTyValGluPheProIleLysProGlyT 53  
:::|||||

312 .....TTCTGAACCCCATACAGTA.....GGA 336  
:::|||||

53 hrThrLeuThrLysGluLysIleGluTyTyValGluTrpAlaLeuasp 69  
:::|||||

337 GCAGGATCAAAACTCCAAATGATGATGACATCACTATGCCTCGAC 386  
:::|||||

```

70 AlaThrAla.....TyrLysGluPhe 76
   :::::|
387 AGCAGTGCMAAAATGAACCCTGGGAGAGTTTATGATAGAAAGTCCGG 436
76 eArgValValGluLeuAsp.....ProSerAlaLysIleGluValP 90
   |||||:|:|:|:|
437 AAGAGTGATGAGTGTAAAAGCATCTGCACCTGCTTCAATTCTACACTT 486
90 hTyrTyrAspLysAsnLysLysGluGluThrLysSerPhePro... 105
   :::::|:|:|:|:|:|:|
487 CCAATTCGCAATAAAGTAAAAAGGAAAGTGATTTATGATATCCGCT 536
106 .....IleThrGluLysGlyPheValValProAspLeuSerGluHt 119
   :::::|:|:|:|:|:|:|
537 CATGCAGCATTCGCTTTGGAAACTCTAGTATTTCCAGACGCTGTGAATCA 586
119 s.....IleLysAsnProGlyPheAsnLeuIleThrL 130
   ||| |||||:|:|:|:|
587 CCTTAATTTTCCATCTCAATGTGTAATCCGGAGAGAGATACGTCATT 636
130 ySValVal 132
   ::|||
637 CAGTGGTG 644

```

seq\_name: gb\_est1:AW485868

seq\_documentation\_block:

```

LOCUS      AW485868                393 bp    mRNA    linear    EST 09-JUL-2000
DEFINITION 68470 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   AW485868
VERSION     AW485868.1  GI:7055974
KEYWORDS    EST.
SOURCE      pig.
ORGANISM    Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE   1 (bases 1 to 393)
            Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.,
            and Keeler,J.W.
            Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
JOURNAL     Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -mismatch 12 options.
PCR PRIMERs
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 28 row: N column: 23
Seq primer: ATTTAGGTGACACTATAG.

```

FEATURES

```

Source      1. 393
            /organism="Sus scrofa"
            /db_xref="taxon:9823"
            /clone_lib="MARC 1P1G"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: PCMV SPORT6; Site_1: xbaI; Site_2: xhoI;
            library made from pooled tissue from day 11, 13, 15, 20,
            and 30 embryos."
BASE COUNT  105 a      83 c      120 g      85 t
ORIGIN

```

alignment\_scores:

```

Quality:      83.00      Length:      116
Ratio:        1.456      Gaps:      5
Percent Similarity: 49.138      Percent Identity: 25.000

```

alignment\_block:

us-09-728-670-10 x AW485868/rev ..

Align seg 1/1 to reverse of: AW485868 from: 1 to: 393

```

20 ProThrGlyProTyrIleuMetValAsnValThrGlyValAspSerLysG1 36
   |||||:|:|:|:|:|:|:|
333 CCAACCAAGGCGCATTTTCTAGCAATGGGGCTACTCGACAAAGCTCTGCTGG 284
36 yAsnGluLeu.....LeuSerProH1sTyrV 45
   |||:|
283 TTCACTTCTCCAGGGAAGAGTCCCAAAGCTGTAGACCTCACCATTCTTCTG 234
45 aGlu.....Phe 47
   ||::|
233 TCAGAAAGACTCCGTTGTTTCACTCCAAAGGCGCTGCGTGAAGACAGTCTTT 184
48 ProIleLysProGly.....ThrThrLeuThrLysGluLysIleGluTyr 62
   |||:|:|:|:|:|:|:|
183 CCTGTCCAGCCTGGCAATCTCTGTGTGTACAGAACAG..... 145
62 rTyrValGluTrpAlaLeuAspAlaThrAlaTyrLysGluPheArgValV 79
   |||:|:|:|
144 .....GATCCTCGCTGCCAGACATGGACAGGCCCTC 114
79 aGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspLysAsn 95
   :::::|:|:|:|:|:|:|:|
113 TTCTTTTGATCCTTCAGACCTCTGTGAGCTTCTCTTTGAGTCCATC 64
96 LysLysLysGluGlu.....ThrLysSerPheProIleThrGluLys 109
   :::::|:|:|:|:|:|:|
63 GGTCAGCGGGAGCAGCTCCATCCAAATCATTTCCCTCTCTCGGAAAGG 16

```

seq\_name: gb\_gss:AZ052631

seq\_documentation\_block:

```

LOCUS      AZ052631                500 bp    DNA    linear    GSS 30-MAR-2000
DEFINITION RPCI-23-413D16.TJ RPCI-23 Mus musculus genomic clone RPCI-23-413D16
ACCESSION   AZ052631
VERSION     AZ052631.1  GI:7343790
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 500)
            Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-23
JOURNAL     Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-413D16.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buhalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buhalo.edu/orderingframe.htm)
            or from Reseach Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tbdb/bac_ends/mouse/bac_end_intro.html
            Plate: 413 row: D column: 16
            Seq primer: SP6
            Class: BAC ends.

```







cocoanut milk (filter sterilized and added after autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8). Fresh medium was added every 7 days, and cultures were grown at 25 C, with 12hrs of light and continuous shaking."

BASE COUNT 192 a 140 c 125 g 165 t  
ORIGIN

alignment\_scores:      Quality: 80.00      Length: 128  
                         Ratio: 1.250      Gaps: 6  
Percent Similarity: 50.000      Percent Identity: 26.562

alignment\_block:  
US-09-728-670-10 x BI203214 ..

Align seg 1/1 to: BI203214 from: 1 to: 622

```
20 ProThrGlyProGlyLeuMetValAsnValThrGlyValAspSerLysG1 36
|||||
270 CCGACAGGGAATTGCCCGCCGTCAGACACACACCTATGAT..... 311
36 YAsnGluLeuSerProHisTyrValGluPheProIleLysProGlyT 53
:::|||||
312 .....TTCCTGAACCCCATACAGTA.....GGA 336
53 hrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAsp 69
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
337 GCAGGATCAAAAATCTCAAAATGATATGACATCAACTATGACATCGAC 386
70 AlaThrAla.....TyrLysGluPhe 76
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
387 AGCAGTGCAAAATGAAACCCGTTGGGAGAGATTATGATAAGAGTCGG 436
76 eArgValAlaGluLeuAsp.....ProSerAlaLysIleGluValT 90
|||||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
437 AAGAGTGATGAGTGAAGACATCTGCACCTGCTGTTCAATCTTACACTT 486
90 hrTyrTyrAspLysAsnLysLysGluGluThrLysSerPhePro... 105
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
487 CCAATTGGGATAAAGTAAAGGAAAGGAGGATTTATGTATCCGCCCT 536
106 .....IleThrGluLysGlyPheValValProAspLeuSerGluH1 119
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
537 CATGCAGCATGTGGCTTGGAAACTCTAGTATTTCCAGACGCTGTGAATCA 586
119 s.....IleLysAsnProGly 124
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
587 CCTTAATTTCCATCTCAATTTGTAATCCGGCA 620
```

seq\_name: gb\_est2:BI769234

seq\_documentation\_block:

LOCUS BI769234 1100 bp mRNA linear EST 25-SEP-2001  
DEFINITION 603060063F1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5209527 5',  
mRNA sequence.

ACCESSION BI769234  
VERSION BI769234.1 GI:15760812

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Life Technologies, Inc.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM1525 row: m column: 16

High quality sequence stop: 739.

#### FEATURES

source

1. 1100

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5209527"

/lab\_host="DH10B"

/note="Organ: pooled lung and spleen; Vector: PCMV-SPORT6;

Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source

anonymous pool of 24 week female lung, 16 week female

spleen, and 20-22 week male spleens. Library is oligo-dT

primed and directionally cloned (EcoRV site is destroyed

upon cloning). Average insert size 1.4 kb, insert size

range 1-3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 026. Note:

this is a NIH\_MGC Library."

BASE COUNT 266 a 329 c 325 g 180 t  
ORIGIN

alignment\_scores:      Quality: 80.00      Length: 123  
                         Ratio: 1.333      Gaps: 5  
Percent Similarity: 48.780      Percent Identity: 26.829

alignment\_block:

US-09-728-670-10 x BI769234 ..

Align seg 1/1 to: BI769234 from: 1 to: 1100

```
38 GluLeuLeuSerProHisTyrValGluPheProIleLysProGlyThrTh 54
|||||
581 GAACCTCTGAGCCCTAAC.....CCAGATGCAGAGGAGGACAAC 618
54 rLeuThrLysGlu.....LysIleGluT 62
|:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
619 GATCGTCGGGAGATTGAGACGCTGTATCACTCCTTAAGAGCATGAGT 668
62 yrTyrValGluTrp.....AlaLeuAspAlaThrAlaTyrLys 74
|||
669 ATCGGCTCTCTTGCCCGCTTGAGGAGCTTAGACTTGGCCATCTACAAT 718
75 Glu.....PheArgValAlaGluLeuAspProSe 84
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
719 AGCAATCAATGTGTCATCACCATTCTCTTGGAAATCTCACAACTCTC 768
84 rAlaLysIleGluValThrTyrTyrAspLysAsnLysLysGluGluT 101
|||||
769 AGCAAGCTTGAACCGCTCACCTTAGAAAGAGAGCAGAGAGGAGCCCA 818
101 hrLysSerPheProIleThrGluLysGlyPheValValProAsp..... 115
|||||
819 CCAAGGAGCTCTCTGACAGACAAATTGGGAGACACACTGAGCCAGGCTT 868
116 .....LeuSerGluHisIleLysAsnProGlyPheAsnLeuIleThrLy 130
|||
869 GAAGCACTCCGGGAATCCGTAAGCCCTTGGGATTCACACGCTTCAGA 918
130 sValValIleGluLysLys 136
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
919 AATCGTCAAGAGAAAGAAA 937
```

seq\_name: gb\_est2:BG947905

seq\_documentation\_block:

LOCUS BG947905 602 bp mRNA linear EST 12-JUN-2001

DEFINITION IPI\_6.G12.bl\_A002 Immature panicle 1 (IPI) Sorghum bicolor cDNA,  
mRNA sequence.  
ACCESSION BG947905  
VERSION BG947905.1 GI:14366094  
KEYWORDS EST.  
SOURCE sorghum.  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoidae; Andropogoneae; Sorghum.  
REFERENCE 1 (bases 1 to 602)  
AUTHORS Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt  
,L.H.  
TITLE An EST database from Sorghum: developing preanthesis panicles  
JOURNAL Unpublished (2001)  
COMMENT Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmprratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for high quality sequence is  
20. Three-prime sequences, which are obtained with PolytMx or T7  
sequencing primer, are presented as the reverse complement.  
Seq primer: JEN REV  
High quality sequence stop: 516  
POLYA=NO.

FEATURES  
source  
1.602  
/organism="Sorghum bicolor"  
/cultivar="Brx623"  
/db\_xref="taxon:4558"  
/clone\_lib="Immature panicle 1 (IPI)"  
/note="Organ: Developing preanthesis panicles; Vector:  
pBluescript II SK(-) from Lambda Zap II; Site\_1: XhoI;  
Site\_2: EcoRI; The library was made from poly-A RNA in the  
cloning vector lambda Zap II. Clones to be sequenced were  
prepared by mass excision."

BASE COUNT 156 a 130 c 167 g 149 t  
ORIGIN

alignment\_scores:  
Quality: 79.00 Length: 109  
Ratio: 1.362 Gaps: 5  
Percent Similarity: 53.211 Percent Identity: 26.606

alignment\_block:  
US-09-728-670-10 x BG947905 ..

Align seg 1/1 to: BG947905 from: 1 to: 602

```
23 proTyrLeuMetValAsnValThrGlyValAspSerLysGlyAsnGluLe 39
|||||  |||  ::|  |||::|::|::|::|::|::|::|::|::|::|
214 CCATGGTGGCTGTCCCTTTTCT.....GACACCGAAGCGCGTGGCGC 257
39 uLeuSer.....ProHisTyrValGluPhep 48
|||||  |||  ::|  |||::|::|::|::|::|::|::|::|::|
258 CCTGATGCCCGGTTCAAGTCTCTGCTATGCCACACCTGCTCATCTTG 307
48 rolleLysProGlyThrThrLeuThrLysGluLysIleGluTyrTyrVal 64
|||||  |||  ::|  |||::|::|::|::|::|::|::|::|::|
308 ATGCAAAACTGGTGAAGTTTACACTGAAGATGGTGTGAATTTGTGAGT 357
65 GluTyrPalaLeuAspAlaThrAlaTyrLysGluPheArgValValGluLe 81
|||||  |||  ::|  |||::|::|::|::|::|::|::|::|::|
358 GAGTATGGCGGTGAGGCTTACCTTCACACAGATGAGATCAATGAACT 407
81 uAspProSerAlaLysIleGluValThrTyrTyrAspLysAsnLysLysL 98
|::|  |||  |||  ::|  |||::|::|::|::|::|::|::|::|
408 GAAGGAACAGAAAGGCAGAA.....AAGGAGAAATCAAACTA 445
```

98 ysGlu.....GluThrLysSerPheProIleThrGluLys 109  
::| ||| ::| |||::|::|::|::|::|::|::|::|::|  
446 TTCAAAGCGTGTGGCACATCTACTGAGACTATCTCATTTCAACAAG 495  
110 GlyPheValValProAspLeuSerGlu 118  
||| ||| ::| |||::|::|::|::|::|::|::|::|::|  
496 GGAGACAAGGTACCC...ATCTCTGAG 519

seq\_name: gb\_est2:BE998438

seq\_documentation\_block:  
LOCUS BE998438 585 bp mRNA linear EST 06-OCT-2000  
DEFINITION EST430161 GVSN Medicago truncatula cDNA clone pgVSN-9P2, mRNA  
sequence.  
ACCESSION BE998438  
VERSION BE998438.1 GI:10698714  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.  
REFERENCE 1 (bases 1 to 585)  
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town  
,C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.  
ESTs from senescent nodules of Medicago truncatula  
JOURNAL Unpublished (2000)  
COMMENT Contact: Carroll P. Vance  
Department of Agronomy and Plant Genetics  
University of Minnesota  
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA  
Tel: 612 625 5715  
Fax: 651-649-5058  
Email: vance004@maroon.tc.umn.edu  
University of Minnesota name: M272416e TIGR sequence name:  
MTKAP85TK More information is available at:  
http://chryslr.tamu.edu/medicago  
Seq primer: SKmod (CTA gAA CTA gTg gAT CC).  
Location/Qualifiers  
1.585  
/organism="Medicago truncatula"  
/cultivar="genotype Al7"  
/db\_xref="taxon:3880"  
/clone="pgVSN-9P2"  
/clone\_lib="GVSN"  
/tissue\_type="senescent root nodules"  
/dev\_stage="mixture of effective nodules from 40 day old  
plants harvested 36 hours post shoot removal and nodules  
collected from 2 month old plants at mid-pod stage"  
/lab\_host="E. coli strain SOLR"  
/note="Vector: pBluescript SK +/-; Site\_1: EcoRI; Site\_2:  
XhoI; cDNA was prepared from polyA+ enriched RNA from the  
mixture of effective nodules of 40 day old plants  
harvested 36 hours post shoot removal and nodules  
collected from 2 month old plants at mid-pod stage. The  
cDNA was directionally ligated into the Uni-ZAP XR vector  
from Stratagene and packaged using Gigapack III Gold  
packaging extracts. Plasmids containing cDNA inserts were  
excised from the recombinant lambda-ZAP phage using  
Ex-Assist helper phage and propagated in SOLR cells."

BASE COUNT 181 a 80 c 137 g 187 t  
ORIGIN

alignment\_scores:  
Quality: 78.50 Length: 157  
Ratio: 0.981 Gaps: 9  
Percent Similarity: 50.955 Percent Identity: 24.204

alignment\_block:  
US-09-728-670-10 x BE998438 ..



Align seg 1/1 to: BE998438 from: 1 to: 585

```
1 SerSerSerPheAspLysGlyLysTyrLysLysGlyAspAlaSerTy 17
|||||:::  ::  |||  ::  |||||
45 TCAAGTGTGTAGAGACATTGAACAAGGGTTATGAAATATCAGTTA 94
17 rPheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAsp 34
|||||  |||:::|||||  ::|||  ::|
95 TTTT...CCTGTAGGACCC.....ATTACACAATAATAGGAT 126
34 erLysGlyAsnGluLeuLeu..... 40
|||||:::|||||:::
127 CAAGTAACAATGATGTTGTTGGTGATGACATGAAATGTTGAATGGTTG 176
41 .....SerProHisTyrValGluPheProIleLysPr 51
|||||:::|||||
177 AAAAACCAACACCAAAATCTGTTTGTATGTTCTTT.....GGAAG 220
51 ogLysThrLeuThrLysGlyLysIleGluTyrTyrValGluTrpAlaL 68
|||||  |||:::|||||:::  ::  ::::|
221 TGGTGGAAACACTTCTCAAGACAAATCAAT...GAGATAGCTTTGGTT 267
68 euAspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProSer 84
||:::  ::::  ::  ::  |||||:::  |||||
268 TGGAAATGAGTGGTCAAGATTCATTGGGTTGTGAGGCA...CCAAGT 314
85 AlAlaLysIleGluValThrTyrTyrAspLysAsnLysLys..... 97
:::|||||:::  |||  :::|||||:::
315 GATTCAAGTCAATGACACCTATCTTGAATCTACAAATGAAGATCCTTGAA 364
98 .....LysGluGluThrLysSerPheProIleThrg 108
|||:::|||||
365 ATTCTTCCAGAGCGTTTCACGAAGACCAAA.....G 399
108 LuLysGlyPheValValPro.....AspLeuSerGlu 118
|||||  |||:::|||||  ::::  :::
400 AAAAAGTTTCATTGTCATCATGCGCACCTCAAGTTGAATACTTAAA 449
119 HisIleLysAsnProGlyPhe 125
|||  ::  |||||
450 CATAGTTCAGTTGGTGGATTT 470
```

seq\_name: gb\_est2:BI953150

seq\_documentation\_block: 481 bp mRNA linear EST 19-OCT-2001  
LOCUS BI953150  
DEFINITION HVSMEM0011F14f Hordeum vulgare green seedling EST library  
HYCDNA0014 (Blumeria infected) Hordeum vulgare cDNA clone  
HVSMEM0011F14f, mRNA sequence.

ACCESSION BI953150  
VERSION BI953150.1 GI:16297584  
KEYWORDS EST.  
SOURCE barley.  
ORGANISM Hordeum vulgare

REFERENCE 1 (bases 1 to 481)  
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,  
Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons  
,J., Oates,R. and Metcal,D.  
TITLE Development of a genetically and physically anchored EST resource  
for barley genomics: Blumeria infected Morex (compatible) seedling  
cDNA library

JOURNAL Unpublished (2001)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu

Total bq bases = 436  
Seq primer: AATTAAACCCCTCACTAAAGGC  
High quality sequence stop: 475.  
location/Qualifiers  
1. 481

FEATURES  
source  
/organism="Hordeum vulgare"  
/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone="HVSMEM0011F14f"  
/clone\_lib="Hordeum vulgare green seedling EST library  
HVCDNA0014 (Blumeria infected)"  
/tissue\_type="green seedling leaf"  
/lab\_host="TJCI21"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Morex (mla) plants were greenhouse grown in the R  
Wise lab at Iowa State University, Ames, IA; 7 day old  
green seedlings were infected with isolate 5874 of  
Blumeria graminis f. sp. hordei, and leaves were harvested  
24, 48 and 72 hr post-inoculation and snap frozen (Wise).  
In the TJ Close lab at the University of California,  
Riverside, total RNA was prepared from each sample pool,  
equal quantities of all three RNA pools were combined,  
poly(A) RNA was purified from the mixture, one primary  
unamplified cDNA library was made, and 1 million pfu were  
in vivo excised to give pBluescript SK(-) cDNA phagemids  
(Chin). Phagemids were plated and picked at the Clemson  
University Genomics Institute (CUGI) (Begum, Palmer,  
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA  
sequencing and sequence analysis were performed at CUGI  
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).  
The sequence has been trimmed to remove vector sequence  
and contains a minimum of 100 bases of phred value 20 or  
above. For more details on library preparation and  
sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders  
Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)\*

BASE COUNT 136 a 81 c 121 g 142 t 1 others

alignment\_scores:  
Quality: 78.00 Length: 116  
Ratio: 1.393 Gaps: 5  
Percent Similarity: 48.276 Percent Identity: 25.862

alignment\_block:  
US-09-728-670-10 x BI953150 ..

Align seg 1/1 to: BI953150 from: 1 to: 481

```
16 SerTyrPheGluProThrGlyProTyrLeuMetValAsnValThrGlyVa 32
:::|||||  |||:::||||  |||
74 GCCTATTGTCAAAGATG...CCGTGGTGGCAGTTCCTTCTCT..... 115
32 lAspSerLysGlyAsnGluLeuLeuSer..... 41
|||||:::|||||:::  |||:::
116 .GACTCTGAAGAGCGTAAACCTTGATGAGCGGTTTGAGGTCAACGGTA 164
42 ..ProHisTyrValGluPheProIleLysProGlyThrThrLeuThrLys 57
|||||  |||  ::  |||  |||  :::|||||
165 TTCACACACCTGTTTCTTCTGATGCAAAAACGTGTAAGTTCTTACTGAT 214
58 GluLysIleGluTyrTyrValGluTrpAlaLeuAspAlaThrAlaTyrLy 74
|||  :::|||||  |||:::|||||  ::  :::
215 GAAGAGTTGAGTTTGAAGTGAATATGCTATAGAAGCTTATCCTTTTAC 264
74 sGluPheArgValValGluLeuAspProSerAlaLysIleGluValThrT 91
|||:::  |||||  :::
265 AACTGAGAGATCAATGAATTTGAAGAACAA..... 295
```

91 yRTYrAspLysAsnLysLysGluGluThr..... 101  
::||| ||| ::::|||  
296 ....GAAAAGCAGCTAAGCATATCAACTATTATAGTGTGCTTGGT 340  
102 .....LysSerPheProIleThrGluLysGlyPheValPro 114  
:.....: |||::: ||| |||||  
341 ACAGCAATCGTCCTAGCTAATTCAACACGCGGAGAGGATGCC 388

seq\_name: gb\_est2:BG581416

seq\_documentation\_block:

LOCUS BG581416 708 bp mRNA linear EST 11-APR-2001  
DEFINITION EST483149 GVN Medicago truncatula cDNA clone pGVN-64F2 5' end, mRNA  
sequence.

ACCESSION BG581416  
VERSION BG581416.1 GI:13596480

KEYWORDS EST.

SOURCE barrel medic.  
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.  
1 (bases 1 to 708)

REFERENCE  
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Town  
C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.  
ESTs from one month old nitrogen-fixing root nodules of Medicago  
truncatula, 2001  
JOURNAL Unpublished (2001)

COMMENT Contact: Carroll P. Vance  
Department of Agronomy and Plant Genetics  
University of Minnesota  
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA  
Tel: 612 625 5715  
Fax: 651-649-5058  
Email: vance004@maroon.tc.umn.edu  
University of Minnesota name: M382475e TIGR sequence name:  
MTCDJ25TK More information is available at: <http://www.medicago.org>  
Seq primer: Skmod (CTA gAA CTA gtg gAT CC).

#### FEATURES

source  
1..708  
Location/Qualifiers

/organism="Medicago truncatula"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone\_lib="pGVN-64F2"  
/clone\_lib="GVN"  
/tissue\_type="N2-fixing root nodules"  
/dev\_stage="effective root nodules harvested one month  
post inoculation with Sinorhizobium meliloti"  
/lab\_host="E. coli strain XLOLR"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; cDNA was prepared from polyA+ enriched RNA from  
effective root nodules harvested one month post  
inoculation with Sinorhizobium meliloti. The cDNA was  
directionally ligated into the Uni-ZAP XR vector from  
stratagene and packaged using GigaPack III Gold packaging  
extracts. Plasmids containing cDNA inserts were excised  
from the recombinant lambda-ZAP phage using Ex-Assist  
helper phage and propagated in XLOLR cells."

BASE COUNT 209 a 150 c 179 g 170 t

ORIGIN

alignment\_scores:

Quality: 78.00 Length: 113  
Ratio: 1.393 Gaps: 1  
Percent similarity: 49.558 Percent identity: 23.894

alignment\_block:

US-09-728-670-10 x BG581416 ..

Align seg 1/1 to: BG581416 from: 1 to: 708

12 GlyAspAlaSerTyrPheGluProThrGlyProTyrLeuMetValAs 28  
||||||| ::| ||| ||||| |||  
345 GGATGATGATGACGGTGTCTTGTCTTCTGATGCTCCGATTCTT..... 386  
28 nValThrGlyValAspSerLysGlyAsnGluLeuLeuSerProHisTyrV 45  
387 .....CCTCCACCTG 396  
45 aGluPheProIleLysProGlyThrThrLeuThrLysGluLysIleGlu 61  
|||::: ::| ||| ::::: ::|  
397 GTGAGATGAGAGTCGAGAGAGGTTACGCTCTTCGCAATGGCCGCTCAA 446  
62 TyrTyrValGluThrPalaLeuAspAlaThrAlaTyrLysGluPheArgVa 78  
::||| ::||| |||||::|||::  
447 AATGCCATTGAGCTAGAGGACAGAGAAAGGAAAGAAATGAGATT 496  
78 lValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspLysA 95  
: ::::: ::| ::| ::|||::: |||::|  
497 GAAGATTATTGAGAGAGGCTGAGGAATATTAAGGTGCTTCTATGAGAAAA 546  
95 snLysLysLysGluGluThrLysSerPheProIleThrGluLysGlyPhe 111  
::||| ::| |||||::: |||::|  
547 GGAGCTTAATGTTGAGACTTAACAAGTTCAAAATAGAGAAAGGAGAA 596  
112 ValValProAspLeuSerGluHisIleLysAsnProGly 124  
||||| ::| |||::: |||  
597 GTTGTCTCGTAGCTAATCAAGAGAAAGTTTCACAAAGAGGCC 635

seq\_name: gb\_est2:BE455199

seq\_documentation\_block:

LOCUS BE455199 711 bp mRNA linear EST 22-OCT-2001  
DEFINITION HVSMeh0096J04f Hordeum vulgare 5-45 DAP spike EST library  
HVCDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMeh0096J04f,  
mRNA sequence.

ACCESSION BE455199  
VERSION BE455199.2 GI:13189831

KEYWORDS EST.

SOURCE barley.  
ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
; Trilicaceae; Hordeum.  
1 (bases 1 to 711)

REFERENCE  
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu  
Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton  
,R.D., Close,S.J., Oates,R. and Main,D.  
Development of a genetically and physically anchored EST resource  
for barley genomics: Morex 5-45 DAP spike cDNA library  
Unpublished (2001)

JOURNAL On Jul 26, 2000 this sequence version replaced gi:9464636.

COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: [rwing@clemson.edu](mailto:rwing@clemson.edu)

Total bp bases = 568  
Seq primer: AATTAACTCTCACTAAAGG  
High quality sequence stop: 704.

#### FEATURES

source

1..711  
Location/Qualifiers  
/organism="Hordeum vulgare"  
/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone="HVSMeh0096J04f"  
/clone\_lib="HVSMeh0096J04f"  
/clone\_lib="Hordeum vulgare 5-45 DAP spike EST library  
HVCDA0009 (5 to 45 DAP)"  
/tissue\_type="5-45 DAP Spike"  
/lab\_host="SOLR"  
/note="Vector: lambdaZAP; site\_1: EcoRI; site\_2: XhoI;

Plants were grown in the greenhouse at the University of California, Riverside (Fenton, SJ Close, TJ Close). Whole spikes with awns trimmed were collected at 5, 10, 15, 20, 30 and 45 DAP (Fenton). Total RNA was prepared from each pool, equal quantities of all six RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Choi) in the TJ Close lab at the University of California, Riverside. Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see

<http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

BASE COUNT 194 a 134 c 186 g 197 t  
ORIGIN

alignment\_scores: Quality: 78.00 Length: 116  
Ratio: 1.393 Gaps: 5  
Percent Similarity: 48.276 Percent Identity: 25.862

alignment\_block:  
US-09-728-670-10 x BE455199 ..

Align seg 1/1 to: BE455199 from: 1 to: 711

```
16 SerTyrPheGluProThrGlyProTyrLeuMetValAsnValThrGlyVa 32
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
211 GCCTATTTGGCAAGATG...CCGTGGTTGGCAGTTCCTTCTCT... 252
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
32 1aspSerLysGlyAsnGluLeuLeuSer..... 41
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
253 .GACTGTGAAGACGTAATAAACCTTGATGAGCGGTTGAGGTCAACGGTA 301
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
42 ..ProHisTyrValGluPheProIleLysProGlyThrThrLeuThrLys 57
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
302 TTCACACACCTGTTCCTTGCATGCAAAAACCTGGAAGTCTTACTGAT 351
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
58 GluLysIleGluTyrTyrValGluTrpAlaLeuAspAlaThrAlaTyrLy 74
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
352 GAAGGAGTTGAGTTGTAGTGAATATGATATAGAGCTTATCCTTTTAC 401
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
74 sGluPheArgValValGluLeuAspProSerAlaLysIleGluValThr 91
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
402 AACTGAGAGGATCAATGAATTGAAGAACAA..... 432
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
91 yTyrAspLysAsnLysLysLysGluGluThr..... 101
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
433 .....GAAAAGGACGTAAGATATCAAACTATTCATAGTGTGCTTGGT 477
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
102 .....LysSerPheProIleThrGluLysGlyPheValValPro 114
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
478 ACAGCAAAATCGTGCTACGTAATTTCAAACACGGGGAAGAGGTACCC 525
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

seq\_name: gb\_est2:BF260862

seq\_documentation\_block:

LOCUS BF260862 733 bp mRNA linear EST 22-OCT-2001  
DEFINITION HVSMEF0023C06f Hordeum vulgare seedling root EST library HVCDNA0007  
(Etiolated and unstressed) Hordeum vulgare cDNA clone  
HVSMEF0023C06f, mRNA sequence.

ACCESSION BF260862  
VERSION BF260862.2 GI:13121015  
KEYWORDS EST.  
SOURCE barley.

ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae ; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 733)  
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton ,R.D., Oates,R. and Main,D.

Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling root cDNA library Unpublished (2001)

COMMENT On Nov 16, 2000 this sequence version replaced gi:11189975.  
Contact: Wing RA

Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293

Email: [rwing@clemson.edu](mailto:rwing@clemson.edu)  
Total hg bases = 555  
Seq primer: AATTAACTTCACCTAAAGCG

High quality sequence stop: 722.  
location/Qualifiers

FEATURES  
source

1..733  
/organism="Hordeum vulgare"

/cultivar="Morex"

/db\_xref="taxon:4513"

/clone="HVSMEF0023C06f"

/clone\_lib="Hordeum vulgare seedling root EST library HVCDNA0007 (Etiolated and unstressed)"

/tissue\_type="Seedling root"

/lab\_host="TUC121"

/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedling roots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates , Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see

<http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001)

Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

BASE COUNT 181 a 196 c 184 g 172 t

ORIGIN

alignment\_scores: Quality: 78.00 Length: 116  
Ratio: 1.393 Gaps: 5  
Percent Similarity: 48.276 Percent Identity: 25.862

alignment\_block:

US-09-728-670-10 x BF260862 ..

Align seg 1/1 to: BF260862 from: 1 to: 733

```

16 SerTyrPheGluProThrGlyProTyrLeuMetValAsnValThrGlyVa 32
   ::::::::::::::::::::
396 GCCTATTTCGCAAGATG...CCGTGTTGGCAGTTCCTTCTCT... 437
   |||||
32 LaspSerLySGlyAsnGluLeuLeuSer..... 41
   |||||
438 .GACTCTGAAGGACGTAACCAACCTTGATGACGGCTTGAGGTCAACGGTA 486
   |||||
42 ..ProHisTyrValGluPheProIleLeysProGlyThrThrLeuThrLys 57
   |||||
487 TTCACACACCTTGTTTTCCTTGATGCAAAAAGTGTGAGTTCCTTACTGAT 536
   |||||
58 GluLysIleGluTyrTyrValGluTrpAlaLeuAspAlaThrAlaTyrLy 74
   |||
537 GAAGAGTTGAGTTGTGAAGTGAATATGCTATAGAAGCTTATCCTTTTAC 586
   |||||
74 sGluPheArgValValGluLeuAspProSerAlaLysIleGluValThrT 91
   |||||
587 AACTGAGAGATCAATGAATTGAAGCAACA..... 617
   |||||
91 YrTyrAspLysAsnLysLysLysGluGluThr..... 101
   ::|||
618 ....GAAAAGGCAGCTAAGGATATCAACTATTCATAGTGTGCTTGGT 662
   ::|||
102 .....LysSerPheProIleThrGluLysGlyPheValPro 114
   ::|||
663 ACAGCAATCGTCCCTACGTAATTTCACACACGCGGAAGAGGTACCC 710
   ::|||

```

seq\_name: gb\_est2:BG583222

seq\_documentation\_block:

LOCUS BG583222 734 bp mRNA linear EST 11-APR-2001  
DEFINITION EST484972 GVN Medicago truncatula cDNA clone pGVN-72E3 5' end, mRNA  
sequence.

ACCESSION BG583222  
VERSION BG583222.1 GI:13598286

KEYWORDS EST.  
SOURCE barrel medic.

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.

1 (bases 1 to 734)

REFERENCE  
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Town  
,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.

TITLE ESTs from one month old nitrogen-fixing root nodules of Medicago  
truncatula, 2001

JOURNAL Unpublished (2001)

COMMENT Contact: Carroll P. Vance

Department of Agronomy and Plant Genetics  
University of Minnesota

411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA

Tel: 612 625 5715

Fax: 651-649-5058

Email: vance004@maroon.tc.umn.edu

University of Minnesota name: M384298e TIGR sequence name:

MTCEE26Tk More information is available at: <http://www.medicago.org>

seq primer: SKmod (CTA gaa CTA gta gAT CC).

location/Qualifiers

1. 734

/organism="Medicago truncatula"

/cultivar="genotype A17"

/db\_xref="taxon:3880"

/clone="pGVN-72E3"

/clone\_lib="GVN"

/tissue\_type="N2-fixing root nodules"

/dev\_stage="effective root nodules harvested one month  
post inoculation with Sinorhizobium meliloti"

/lab\_host="E. coli strain XL0R"

/note="Vector: pBluescript SK-; site\_1: EcoRI; site\_2:

XhoI; cDNA was prepared from polyA+ enriched RNA from  
effective root nodules harvested one month post  
inoculation with Sinorhizobium meliloti. The cDNA was  
directionally ligated into the Uni-ZAP XR vector from  
stratagene and packaged using Gigapack III Gold packaging  
extracts. Plasmids containing cDNA inserts were excised  
from the recombinant lambda-ZAP phage using Ex-Assist  
helper phage and propagated in XL0R cells."

BASE COUNT 221 a 159 c 185 g 169 t  
ORIGIN

alignment\_scores:

Quality:	78.00	Length:	113
Ratio:	1.393	Gaps:	1
Percent Similarity:	49.558	Percent Identity:	23.894

alignment\_block:

US-09-728-670-10 x BG583222 ..

Align seg 1/1 to: BG583222 from: 1 to: 734

```

12 GlyAspAspAlaSerTyrPheGluProThrGlyProTyrLeuMetValAs 28
   |||||
307 GGTGATGATGACGGGTGTTTGTCTTGATGCTGCGATTCTT..... 348
   |||||
28 nValThrGlyValAspSerLySGlyAsnGluLeuLeuSerProHisTyrV 45
   |||||
349 .....CCTCCACCTG 358
   |||||
45 alGluPheProIleLysProGlyThrThrLeuThrLysGluLysIleGlu 61
   |||||
359 GTGAGATGAGTCGAGAGAGCTTACGCTCTTCGCGAATGCGCGCTCAA 408
   |||||
62 TyrTyrValGluTrpAlaLeuAspAlaThrAlaTyrLysGluPheArgVa 78
   ::|||
409 AATGCCATTGAGCTAGAGAGAGAAGAGAAAAGGAAAAGAAATGAGATT 458
   ::|||
78 lValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspLysA 95
   ::|||
459 GAGATTATTGAGAGAGGCTGAGGAATATAGGTGCTTCTATGAGAAAA 508
   ::|||
95 snLysLysLysGluGluThrLysSerPheProIleThrGluLysGlyPhe 111
   ::|||
509 GGAAGCTTATGTTGAGACTAACCAAGGTTCAAAATAGAGAAAGGGAGAA 558
   ::|||
112 ValValProAspLeuSerGluHisIleLysAsnProGly 124
   |||||
559 GTTGTTCGTAGCTAATCAAGAGAAGTTTCACCAAGAGGC 597
   |||||

```

seq\_name: gb\_gss:CNS072XD

seq\_documentation\_block:

LOCUS CNS072XD 1006 bp DNA linear GSS 07-JUL-2001  
DEFINITION clone BA0AB015B05 of library BA0AB from strain CUB 210 of  
Kluyveromyces lactis, genomic survey sequence.

ACCESSION AL426791.1 GI:12209985  
VERSION AL426791

KEYWORDS GSS.

SOURCE Kluyveromyces lactis.

ORGANISM Kluyveromyces lactis

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

1 (bases 1 to 1006)

REFERENCE  
AUTHORS Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,  
Duchateau-Nguyen,G., Lemaire,M., Marmelisse,R., Montrocher,R.,  
Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.

TITLE Genomic exploration of the hemiascomycetous yeasts: 11.

Kluyveromyces lactis

JOURNAL FEBS Lett. 487 (1), 66-70 (2000)

MEDLINE 20584721

REFERENCE 2 (bases 1 to 1006)

AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,



Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,  
 de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,  
 Malpertuy, A., Neveuglise, C., Ozier-Kalogeropoulos, O., Potier, S.,  
 Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,  
 Wincker, P. and Weissenbach, J.  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 FEBS Lett. 487 (1), 3-12 (2000)

**TITLE**  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
**JOURNAL**  
 FEBS Lett. 487 (1), 3-12 (2000)  
**MEDLINE**  
 20584711  
**REFERENCE**  
 3 (bases 1 to 1006)  
**AUTHORS**  
 Genoscope.  
**JOURNAL**  
 Direct Submission  
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This GSS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

**FEATURES**  
 Location/Qualifiers  
 1..1006  
 /organism="Kluyveromyces lactis"  
 /strain="CLIB 210"  
 /variety="lactis"  
 /db\_xref="taxon:28985"  
 /clone="BA0AB015B05"  
 /clone\_lib="BA0AB"

**BASE COUNT** 381 a 179 c 224 g 220 t 2 others  
**ORIGIN**

alignment\_scores:      Quality: 78.00      Length: 125  
                          Ratio: 1.164      Gaps: 4  
                          Percent Similarity: 53.600      Percent Identity: 25.600

alignment\_block:  
 US-09-728-670-10 x CNS072XD ..

Align seg 1/1 to: CNS072XD from: 1 to: 1006

```

6  LysGlyLysTyrLysGlyAspAspAlaSerTyrPheGluProThrG1 22
   ||| ||| ::| |||::|::|::| ||| |||
   ||| ||| ::| |||::|::|::| ||| |||
335 AAGGTCAAACTCAATGATTAAGATGAAGCAAAACCTTCAGAGAGACTGA 384
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
22 yProTyrLeuMetValAsnValThrGlyValAspSerLysGlyAsnGlu 39
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
385 TAAGTTCATTAACAGAGAAATGATACA..GCAATGAAGAGGCCAATTCCG 431
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
39 euleuSerProHisTyrValGluPheProIleLysProGlyThrThrleu 55
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
432 TAGAGAGAGACTAAATAATTAATACTGCTGAGACAACGAAACAGCGTA 481
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
56 ThrLysGluLysIleGluTyrTyrValGluTyrPalaLeuAspAlaThrAl 72
   ||| ||| |||::|::|::|::|::|::|::|::|::|::|::|::|::|
482 ACAAAAGAGAGAA.....GATACAAATTA 504
   ||| ||| |||::|::|::|::|::|::|::|::|::|::|::|::|::|
72 aTyrLysGluPheArgValAlaGluLeuAspProSerAlaLysIleGluV 89
   ||| ||| ::|::|::|::|::|::|::|::|::|::|::|::|::|
505 TAAAAAAGAGAGAGAGCCAGAAAGATGAGGGCCCAAGCAGAGAGA 554
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
89 aLThrTyrTyrAspLysAsnLysLysGluGluThrLysSerPhePro 105
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
555 TA.....GAAGGAGAGAAACAGAACCTTAACAGAAAAAGAC 592
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
106 lIeThrGluLysGlyPheValValProAspLeuSerGluHisIleLysAs 122
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
  
```

593 GTGCAAAACCATCGTTCACATTGGCCCAATCTTCACAGTTTGTTCt.. 640  
 122 nProGlyPheAsnLeuIleThrLys 130  
 641 ...GGCTTTGAGTGGCAACTTAA 661

seq\_name: gb\_gss:CNS073BU

seq\_documentation\_block:  
 LOCUS CNS073BU 1030 bp DNA linear GSS 07-JUL-2001  
 DEFINITION clone BA0AB018D09 of library BA0AB from strain CLIB 210 of  
 Kluyveromyces lactis, genomic survey sequence.  
 ACCESSION AL427312  
 VERSION AL427312  
 AL427312.1 GI:12210506  
 KEYWORDS GSS.  
 SOURCE Kluyveromyces lactis.  
 ORGANISM Kluyveromyces lactis.  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

**REFERENCE**  
 Bolotin-Fukuhara, M., Toffano-Nioche, C., Artiguenave, F.,  
 1 (bases 1 to 1030)  
 Duchateau-Nguyen, G., Lemaire, M., Marmelisse, R., Montrocher, R.,  
 Robert, C., Termier, M., Wincker, P. and Wesolowski-Louvel, M.  
 Genomic exploration of the hemiascomycetous yeasts: 11.  
 Kluyveromyces lactis  
 FEBS Lett. 487 (1), 66-70 (2000)

**TITLE**  
 Genomic exploration of the hemiascomycetous yeasts: 11.  
 Kluyveromyces lactis  
**JOURNAL**  
 FEBS Lett. 487 (1), 66-70 (2000)  
**MEDLINE**  
 20584721  
**REFERENCE**  
 2 (bases 1 to 1030)  
 Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,  
 Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,  
 de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,  
 Malpertuy, A., Neveuglise, C., Ozier-Kalogeropoulos, O., Potier, S.,  
 Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,  
 Wincker, P. and Weissenbach, J.  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 FEBS Lett. 487 (1), 3-12 (2000)

**TITLE**  
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
**JOURNAL**  
 FEBS Lett. 487 (1), 3-12 (2000)  
**MEDLINE**  
 20584711  
**REFERENCE**  
 3 (bases 1 to 1030)  
**AUTHORS**  
 Genoscope.  
**JOURNAL**  
 Direct Submission  
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This GSS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

**FEATURES**  
 Location/Qualifiers  
 1..1030  
 /organism="Kluyveromyces lactis"  
 /strain="CLIB 210"  
 /variety="lactis"  
 /db\_xref="taxon:28985"  
 /clone="BA0AB018D09"  
 /clone\_lib="BA0AB"

misc\_feature  
 complement(<2..>1018)  
 /note="similar to Saccharomyces cerevisiae ORF YUL130c [  
 URA2 : multifunctional pyrimidine biosynthesis protein ]"  
 /evidence="not\_experimental"

**BASE COUNT** 326 a 217 c 198 g 286 t 3 others  
**ORIGIN**

alignment\_scores:      Quality: 78.00      Length: 92  
                          Ratio: 1.472      Gaps: 4

Percent Similarity: 57.609 Percent Identity: 27.174

# alignment\_block:

US-09-728-670-10 x CNS073BU/rev ..

Align seg 1/1 to reverse of: CNS073BU from: 1 to: 1030

```

39 LeuLeuSerProHisTyrValGluPheProIleLysProGlyThrThrLe 55
   |||:: ||| ||||| ::|||::
574 TTAGTCAGACCTTCCTACGCTTATTCAGGTCGCCGAATGAAATACAGTCTA 525
   ::|||:: ::||| |||||:: |||::|||::
55 uThrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAspAlaThra 72
   ::|||:: ::||| |||||:: |||::|||::
524 TTCTAAGATGACTTGGAATCTTATTGAACACAGGCCGTGGAAGTTTCT. 476
   ::|||:: ::||| |||||:: |||::|||::
72 lAtYrLysGluPheArgValValGluLeuAspProSerAlaLysIleGlu 88
   ::|||:: ::||| |||||:: |||::|||::
475 ....CGTGAATATCCAGTTGTT..... 458
   ::|||:: ::||| |||||:: |||::|||::
89 ValThrTyrTyrAspLysAsnLysLysGluGluThrLysSerPhePr 105
   ::||| ||| ::||| |||||:: |||::
457 ATCACCAGATATATTGAAATGCTAAGAAATCGAAATGACGCAAGTTGC 408
   ::|||:: ::||| |||||:: |||::|||::
105 oIleThrGluLysGlyPheValValProAspLysSerGluHisIleLysA 122
   ::|||:: ::||| |||||:: |||::|||::
407 CTTA...GACGCGTGAATTGATTCATGTCGTTTCTGAGCATGTTGAAA 361
   ::|||:: ::||| |||||:: |||::|||::
122 snProGly.PheAsnLeuIleThr 129
   || ||| |||::|||::|||
360 ACGCAGGTGTTCACTCTGTGACG 337

```

seq\_name: gb\_gss:BH483792

## seq\_documentation\_block:

LOCUS BH483792 815 bp DNA linear GSS 13-DEC-2001  
 DEFINITION BOGWK46TR BOGW Brassica oleracea genomic clone BOGWK46, DNA  
 sequence.  
 ACCESSION BH483792  
 VERSION BH483792.1 GI:17691896  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea.  
 ORGANISM Brassica oleracea.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
 REFERENCE 1 (bases 1 to 815)  
 TOWN,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
 Whole genome shotgun sequencing of Brassica oleracea  
 Unpublished (2001)  
 JOURNAL Other\_GSSs: BOGWK46TF  
 COMMENT Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.

## FEATURES

source location/Qualifiers  
 1..815  
 /organism="Brassica oleracea"  
 /strain="TO1000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOGWK46"  
 /clone\_lib="BOGW"  
 /note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared  
 genomic DNA inserted into PHOS1 using BstXI linkers"  
 BASE COUNT 261 a 145 c 159 g 250 t  
 ORIGIN

alignment\_scores:

Quality: 77.50 Length: 66  
 Ratio: 1.890 Gaps: 3  
 Percent Similarity: 62.121 Percent Identity: 28.788

# alignment\_block:

US-09-728-670-10 x BH483792 ..

Align seg 1/1 to: BH483792 from: 1 to: 815

```

5 AspLysGlyLysTyrLysLysGlyAspAspAlaSerTyrPheGluProth 21
   ::||| ||| ||||| ||||| |||||
614 GAGAAAAAATAAGACTAAAGAGAT..... 640
   ::||| ||| ||||| ||||| |||||
21 rGlyProTyrLeuMetValAsnValThrGlyValAspSerLysGlyAsnG 38
   ::|||:: ::||| |||||:: |||::|||::
641 .....TTAACCATCGCCGCGTGCCAGTGGCAACGAGACA 674
   ::|||:: ::||| |||||:: |||::|||::
38 luleLeuSerProHisTyrValGlu.PheProIleLysProGlyThrTh 54
   ::|||:: |||||:: |||||:: |||||::
675 GTATGCTCGCCATATATCTGCTCATTCGATTAACCCCTGGTGTCTC 724
   ::|||:: |||||:: |||||:: |||||::
54 rLeuThrLysGluLysIle.....GluTyrTyrValGluTrpAla 67
   ::|||:: |||||:: |||||:: |||||::
725 GATCTCTCCAGTCAAGTTGACAGAGAAACTATGACAGAGTGTCTCA 770
   ::|||:: |||||:: |||||:: |||||::

```

seq\_name: gb\_est2:BE517562

## seq\_documentation\_block:

LOCUS BE517562 511 bp mRNA linear EST 08-AUG-2000  
 DEFINITION WHE0628\_F09\_L18ZA Wheat ABA-treated embryo cDNA library Triticum  
 aestivum cDNA clone WHE0628\_F09\_L18, mRNA sequence.  
 ACCESSION BE517562  
 VERSION BE517562.1 GI:9741592  
 KEYWORDS EST.  
 SOURCE bread wheat.  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
 ; Triticaceae; Triticum.  
 REFERENCE 1 (bases 1 to 511)  
 Anderson,O.D., Chao,S., Han,P.S., Hsia,C.C., Johnson,R.R., Kang,Y.,  
 Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Tong,J.C., Verhey  
 S.D. and Walker-Simmons,M.K.  
 The structure and function of the expressed portion of the wheat  
 genomes - ABA-treated embryo library  
 Unpublished (2000)  
 JOURNAL Contact: Olin Anderson  
 COMMENT US Department of Agriculture, Agriculture Research Service, Pacific  
 West Area, Western Regional Research Center  
 800 Buchanan Street, Albany, CA 94710, USA  
 Tel: 5105595773  
 Fax: 5105595818  
 Email: oanderson@pw.usda.gov  
 Sequence have been trimmed to remove vector sequence and low  
 quality sequence with phred score less than 20  
 Seq primer: Clontech Matchmaker 3' AD primer.

## FEATURES

source location/Qualifiers  
 1..511  
 /organism="Triticum aestivum"  
 /cultivar="Brevor (soft, white, winter, common wheat)"  
 /db\_xref="taxon:4565"  
 /clone="WHE0628\_F09\_L18"  
 /clone\_lib="Wheat ABA-treated embryo cDNA library"  
 /tissue\_type="Seed embryo"  
 /dev\_stage="Mature dormant seeds"  
 /lab\_host="E. coli DH12S"  
 /note="Vector: pGAD10; Site\_1: EcoRI; Site\_2: XhoI;  
 Embryos were cut from mature, dormant seeds and imbedded in  
 25 microm ABA (abscisic acid) in 5 mM Mes buffer, pH 5.7,  
 for 12 hr at 22 C. The tissue, total RNA, and poly(A)  
 RNA were prepared by Steven Verhey in M.K.  
 Walker-Simmons's lab (USDA-ARS, Washington State Univ.,  
 Pullman, Washington 99164-6420. A cDNA library was made

by Clontech using a combination of random and oligo dt primers. Library was plated and archived by Russell Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 129 a 130 c 114 g 138 t  
ORIGIN

alignment\_scores:                   Quality:       76.50       Length:       79  
                          Ratio:       1.821       Gaps:       3  
Percent Similarity:   53.165       Percent Identity: 30.380

alignment\_block:  
US-09-728-670-10 x BE517562/rev ..

Align seg 1/1 to reverse of: BE517562 from: 1 to: 511

```
16 SerTyrPheGluProThrGlyProTyrLeuMetValAsnValThrGlyVa 32
   ::::| | | | | | | | | | | | | | | | | | | | | | | |
283 GCCTATTGCAAGATG...CCGTGTTGGCAGTTCCTTCTCC..... 242

32 LAspSerLysGlyAsnGluLeuLeuSer..... 41
   ||| | | | | | | | | | | | | | | | | | | | | |
241 .GACTCTGAAGCGCGTAAAGCCTTGATGAGCGGTTTGAGGTCAACGCTA 193

42 ..ProHisTyrValGluPheProIleLysProGlyThrThrLeuThrLys 57
   ||| | | | | | | | | | | | | | | | | | | | | |
192 TTCACACACCTGTATCTTGATGCAAAAACCTGCGAAGTCTTACTGAT 143

58 GluLysIleGluTyrTyrValGluTPrAlaLeuAspAlaThrAlaTyrLy 74
   ||| | | | | | | | | | | | | | | | | | | | | |
142 GAAGGAGTGTAGTTGTGAGTGAATATGGAATGAAGCTTATCTTTTAC 93

74 sGluPheArgValValGluLeuAspProSerAlaLys 86
   ||| | | | | | | | | | | | | | | | | | | | | |
92 AACTGAGAGATCAATGATTTGAAGGAACAAGAAAG 56
```

seq\_name: gb\_gss:AQ501650

seq\_documentation\_block:

LOCUS AQ501650 447 bp DNA linear GSS 29-APR-1999  
DEFINITION V17E9 mTn-3xHA/lacZ Insertion library Saccharomyces cerevisiae  
genomic 5', DNA sequence.

ACCESSION AQ501650  
VERSION AQ501650.1 GI:4707300

KEYWORDS GSS.  
SOURCE baker's yeast.

ORGANISM Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 447)  
Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,  
desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniatis,D., Jansen,R.,  
Umansky,L., Heidtman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,  
Hager,K., Miller,P., Roeder,G.S. and Snyder,M.

TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and  
Gene Disruption

JOURNAL Unpublished (1999)

COMMENT Contact: Kumar A  
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology  
Yale University  
P.O. Box 208103, New Haven, CT 06520-8103, USA  
Tel: 203 432 9949  
Fax: 203 432 6161  
Email: anuj.kumar@yale.edu

te of mTn-3xHA/lacZ insertion.  
Seq primer: GGCTTCTTCTTGTGAAGTAC  
Class: transposon-tagged.

FEATURES  
Source Location/Qualifiers  
1..447  
/organism="Saccharomyces cerevisiae"

/db\_xref="taxon:4932"  
/clone\_lib="mTn-3xHA/lacZ Insertion Library"  
/lab\_host="E. coli"  
/note="Vector: pHSS6-Sal; A yeast genomic DNA library  
(lacking mitochondrial DNA) was prepared in pHSS6-Sal;  
genomic DNA was size-fractionated (DNA of roughly 2-3 kb  
in length) prior to cloning. This library was  
subsequently mutagenized with a mTn-3xHA/lacZ  
minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 138 a 108 c 78 g 122 t 1 others  
ORIGIN

alignment\_scores:                   Quality:       76.00       Length:       127  
                          Ratio:       1.310       Gaps:       6  
Percent Similarity:   45.669       Percent Identity: 22.047

alignment\_block:  
US-09-728-670-10 x AQ501650/rev ..

Align seg 1/1 to reverse of: AQ501650 from: 1 to: 447

```
4 PheAspLysGlyLysTyrLysLysGlyAspAlaSerTyrPheGluPr 20
   ||| | | | | | | | | | | | | | | | | | | | | |
284 TTTACTCCGCGAAGATTTCAGTCGCGCAT..... 255

20 oThrGlyProTyrLeuMetValAsnValThrGlyValAspSerLysGly. 36
   ::::| | | | | | | | | | | | | | | | | | | | | |
254 .....CATATTTCCGGTGTGATTTTACGGTT 227

37 .....AsnGluLeuLeuSerProHisTyrValGluPheProIle 49
   ||| | | | | | | | | | | | | | | | | | | | | |
226 CCGTTTGTGAGAAATTCGTAATTTCAAGCCATAGATTTTTCACCA 177

50 LysProGlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTr 66
   ||| | | | | | | | | | | | | | | | | | | | | |
176 AGATCAAGAGGTACAATCACT.....TG 154

66 PalAlaLeuAspAlaThrAlaTyr.....LysGluPheArgValValG 80
   ||| | | | | | | | | | | | | | | | | | | | | |
153 GATTCCTCCAGCTGCTGAGTACACTTTGGATGAGAAGATTTGCAAGTTG 104

80 LuLeuAspProSerAlaLysIleGluValThrTyrTyrAspLysAsnLys 96
   ||| | | | | | | | | | | | | | | | | | | | | |
103 AATTGTGAT...GGCAAGAAGTTGATTTCACTTTTAC..... 69

97 LysLysGluGluThrLysSerPheProIleThrGluLysGlyPheValVa 113
   ::::| | | | | | | | | | | | | | | | | | | | | |
68 .....CATACTTGCCCTGTTGTGTTCGGGTTTGACGCT 34

113 lProAspLeuSerGluHisIleLysAsnPro 123
   : ::::| | | | | | | | | | | | | | | | | | | | | |
33 CAGTGAACGAAAACTCAAGTTAAGGCCCA 3
```

seq\_name: gb\_est1:AA840229

seq\_documentation\_block:

LOCUS AA840229 453 bp mRNA linear EST 27-FEB-1998  
DEFINITION vw89a06.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone  
IMAGE:1262098 5' similar to gb:D10214 Mouse mRNA for prolactin  
receptor (MOUSE);, mRNA sequence.

ACCESSION AA840229  
VERSION AA840229.1 GI:2915888

KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 453)  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Marras,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

TITLE Theetang,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 JOURNAL The Washu-HMI Mouse EST Project  
 COMMENT Unpublished (1996)  
 Contact: Marra M/Mouse EST Project  
 Washu-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:664650

Seq primer: -28m13 rev1 ET from Amersham  
 High quality sequence stop: 420.

# FEATURES

source  
 1..453  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1262098"  
 /clone\_lib="Stratagene mouse skin (#937313)"  
 /sex="females"  
 /tissue\_type="whole skin"  
 /dev\_stage="11 weeks old"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: skin; Vector: pBluescript SK-; Site\_1: EcoRI  
 ; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
 dt. Whole skin from 11 week old C57BL/6 female mice.  
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'  
 adaptor sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor  
 sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"

BASE COUNT 110 a 129 c 108 g 106 t  
 ORIGIN

## alignment\_scores:

Quality: 76.00 length: 120  
 Ratio: 1.382 Gaps: 6  
 Percent Similarity: 45.833 Percent Identity: 25.000

## alignment\_block:

US-09-728-670-10 x AA840229 ..

Align seg 1/1 to: AA840229 from: 1 to: 453

```

18 PheGluProThrGlyProTyr.....LeuMetValAsnValThrGlyVa 32
   ||| |||||::: ||::|::|::|
164 TTTCCCCCACTTCTGACTGTGAGGACTTGTGCTGTGAGTCTTGAAGT 213
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
32 LAspSerLyseGlyAsnGluLeuLeuSerPro..HisTyrValGluPheP 48
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
214 GGATGACAATGAGGACGAGCGGCTAATGCCATCCCATTCGAAGAGTATC 263
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
48 ro.....IleLysProGlyThrThrLeuThrLysGluLysIleGlu 61
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
264 CGGGTCAAGGTGTTAAACC..... 283
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
62 TyrTyrValGluTTPalaLeuAspAlaThrAlaTyrLysGluPheArgVa 78
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
283 ..... 283
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
78 lValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspLys. 94
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
284 .ACACACCTAGATCCTGACAGTGTGTCATGGAAGCTATGACAGCC 332
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
95 .....AsnLysLysLysGluGluThrLysSerPheProIleThr 107
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
333 ATTCTCTTTGTCTGAAGAAGTGTGAGGAGCCCAAGCCTACCC..... 376
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
108 GluLysGlyPheValValProAspLeuSerGluHisIleLysAsnProGl 124
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

377 ...CTGCCTTCCACATCCCTGAGATCAGTGAAGCAGAGAAATCCTGA 423  
 124 yPheAsnLeu 127  
 |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
 424 GGCAAAATAT 433

seq\_name: gb\_gss:AZ672674

seq\_documentation\_block:

LOCUS AZ672674 885 bp DNA linear GSS 14-DEC-2000

DEFINITION ENT1T05TR Entamoeba histolytica sheared DNA Entamoeba histolytica

ACCESSION AZ672674 genomic, DNA sequence.

VERSION AZ672674.1 GI:11809820

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica

REFERENCE 1 (bases 1 to 885)

AUTHORS Loftus,B., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library

COMMENT Unpublished (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: b.loftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Reverse

Class: shotgun

High quality sequence start: 54

High quality sequence stop: 758.

Location/Qualifiers

1..885

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db\_xref="taxon:5759"

/clone\_lib="Entamoeba histolytica sheared DNA"

/note="Vector: pHO51; Site\_1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + l method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barell, Oxford University Press, 1999)."

BASE COUNT 410 a 66 c 182 g 227 t  
 ORIGIN

## alignment\_scores:

Quality: 76.00 length: 118  
 Ratio: 1.134 Gaps: 7  
 Percent Similarity: 56.780 Percent Identity: 34.746

## alignment\_block:

US-09-728-670-10 x AZ672674 ..

Align seg 1/1 to: AZ672674 from: 1 to: 885

```

19 GluProThrGlyProTyrLeuMetValAsnValThrGlyValAspSerly 35
   ||| |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
94 GAAGTTGAAGAAATTAATCTTCAAGTTGAAGTGAATCTGAAGAAACAA 143
   ||| |||||::|::|::|::|::|::|::|::|::|::|::|::|::|

```



```

35 sglYasnGluleUeSerrProHISTyrrValGluPhePrOIlelYsProc 52  
| :||| |:: ||:::||  
144 A..AAGAATTAATAAGAAAAAGAAAGCAAAATTGAATTTAAAG... 186  
  
52 lYThrThrLeuthrLysGLulsileGUtyTyrrValGIutrpAlaleu 68  
||||| | |||| |  
187 .....AAGAAAA.....GATGTAGAAAAAGCTAAA 213  
  
69 AspalaTrAlatyrLySGluPheargvalValGluleuaspproSerai 85  
|| || ::::|| ::||:::  
214 AGAGCAAAGAAtATtTGAAttTCGT....CATttATCaaAAATgtGC 257  
  
85 alYSilEgluVatlThrtYTYrrAsplysaSnlyslYsgLUglUtHrl 102  
||||:|||||::: |:|::| |  
258 TAAATTAGAgTGcA.....CAAAACAaaaaaatgatTTTTtgTTA 301  
  
102 ysserPheprOieIethrgLUlsGLypHevaIValProaspLeusergLu 118  
|:::||||| || |:: |||||  
302 AAGGaTTTCctaTaTACTGGaag.....attAagaTTtaaaaGaa 342  
  
119 HIsIlELysASnProcLYpheasnlEuILetHrLysVaIVallEGluly 135  
::::: ::|| | ||::|  
343 AGGTTgATgaCTTGtGCTaTTAAAGCaact..ATTTTaTGGAaa 389  
  
135 slYs 136  
|||  
390 AAAA 393  
  
seq_name: gb_gss:BHI58255
```

seq\_documentation\_block:

	LOCUS	BHI58255	890 bp	DNA linear	GSS 24-SEP-2001
DEFINITION	ENTRT14TF Entamoeba histolytica Sheared DNA	Entamoeba histolytica genomic, DNA sequence.			
ACCESSION	BHI58255	BHI58255.1 GI:15731693			
VERSION	BHI58255.1	GI:15731693			
KEYWORDS	GSS.				
SOURCE ORGANISM	Entamoeba histolytica. Entamoeba histolytica Eukaryota; Entamoebidae; Entamoeba.				
REFERENCE AUTHORS TITLE	Loftus,B., Wang,Z., Van Aken,S. and Fraser,C. Determination of clone end sequences from Entamoeba histolytica HM1:IMMS sheared DNA library (2001) Unpublished (2001)				
JOURNAL COMMENT	Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: bjl@fugate.rig.org Clones are derived from the Entamoeba histolytica HM1:IMMS sheared DNA library Seq primer: M13-Forward Class: shotgun High quality sequence start: 10 High quality sequence stop: 559.				
FEATURES source	Location/Qualifiers 1..890 /organism="Entamoeba histolytica" /strain="HM1:IMMS" /db_xref="taxon:5759" /clone_lib="Entamoeba histolytica Sheared DNA" /note="Vector: PHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is detailed in Smith,				

H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrel, Oxford University Press, 1999)."

BASE COUNT 225 a 151 c 92 g 422 t

ORIGIN

alignment\_scores:

	Quality:	76.00	Length:	118
	Ratio:	1.134	Gaps:	7
Percent Similarity:	56.780	Percent Identity:	34.746	

alignment\_block:

US-09-728-670-10 x BH158255/rev ..

Align seg 1/1 to reverse of: BH158255 from: 1 to: 890

```

19 GIUProThrglyProTyrleuMetValAsnValThrglyValAspSerly 35
   ||| ||| |||::|||::||| ||| :::::|
523 GAAGTTGAAGAAAAATACTTCACGTTGAAGTAGATTCGAAGAAACAA 474
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
473 A...AAGAATTAAAAAGAAAAAAGAAGAAGAATGAATTAAG.... 431
   |:::|||||::: |||::: |||:::
35 sclyAsngluLeuSerProHisTyrValGluPheProIleLysProg 52
   |:::|||||::: |||::: |||:::
52 lyThrThrleuThrLysGluLysIleGluTyrTyrValGluTrpAlaLeu 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
430 .....AAGAAGAAA.....GATGTAGAAAAAGCTAAA 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
69 AspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProSerAl 85
   ||| ||| |||::|||::|||::|||::|||::|||::|||::|||
403 AGAGCAAAAGAATATATGTGAATTCGT....CATTTATCAAAATGTGTC 360
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
85 alySIleGluValThrTyrTyrAspLysAsnLysLysLysGluGluThrL 102
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
359 TAAATTAGAAGTTGCA....CAAAACAAAAAAATGATTTTGTGTTA 316
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
102 ySerPheProIleThrGluLysGlyPheValValProAspLeuSerGlu 118
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
315 AAGGATTTCTAATACTTGGAAG.....ATTAAAGATTAAAGAA 275
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
119 HisIleLysAsnProGlyPheAsnLeuIleThrLysValValIleGluL 135
   ::::: ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274 AGGTTGATGACTTGGTGCTATTAAAGCAACT..ATTTAATGGAAGA 228
   ::::: ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
135 slys 136
   |||
227 AAAA 224

```

seq\_name: gb\_est2:BE536273

seq\_documentation\_block:

LOCUS BE536273 822 bp mRNA linear EST 09-AUG-2000

DEFINITION 601062663F1 NIH\_MGC\_10 Homo sapiens cDNA clone IMAGE:3449271 5',

mrna sequence.

ACCESSION BE536273

VERSION BE536273.1 GI:9764918

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 822)

NIH-MGC <http://mgc.ncl.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Incyte Genomics, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://Image.llnl.gov

Plate: LHAM8425 row: m column: 16

High quality sequence stop: 472.

Location/Qualifiers

1..822

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3449271"

/clone\_lib="NIH\_MGC\_10"

/cell\_line="MGC36"

/lab\_host="DH10B"

/note="Organ: cervix; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.5 kb. Library prepared by Life

Technologies."

BASE COUNT 119 a 213 c 287 g 203 t

ORIGIN

#### alignment\_scores:

Quality:	75.50	length:	148
Ratio:	1.180	Gaps:	5
Percent Similarity:	43.243	Percent Identity:	22.297

#### alignment\_block:

US-09-728-670-10 x BE536273/rev ..

Align seg 1/1 to reverse of: BE536273 from: 1 to: 822

```
22 GLYPROTYRLEUMETVALASNVALTHRGLVAL..... 32
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|
643 GGGCCCCACGTTCTTAGGCAATGAATGGGTGCTCCACCCACACATTGGG 594
33 .....AspSerLysGlyAsnGluLeuLeuSerPro.... 42
593 CGCCGTGTGCCCAACCAAGTGTGGGGGCCACTCTTAGCCCGAGACT 544
43 .....HistyValGluPheProIleLys 50
543 CCAGAGCTGGCAAGAGGTCGCMAACACCTATACCGGGCCAAAGAGG 494
51 ProGlyThrThrLeuThrLysGluLysIleGlu..... 61
493 CTTGGGCATAAAAAGCAAGGAGAACTGAGTCTGAACCTCCCTCCC 444
62 .....TyrTyrV 64
443 TCCCCACCCCTGGGTGCATAGACGAGAAAAAATAACGTCCCTCTGGC 394
64 aLGLUTRALaleuaspalaThrAlaTyrLysGluPheArgValValGlu 80
393 TCCCAAGGCGCGCCCGAGCCCGCCGCCACCATCAGTTATATATCCC 344
81 leuAspProSerAlaLysIleGluValThrTyrTyrAspLysAsnLysLy 97
343 CGGACCCCGTCCCGCCCTTGAGGTTTAAAAAAGAGTATGTTTCAATG 294
97 aLysGluGluThrLysSerPheProIleThrGluLysGlyPheValValp 114
293 AAAAGACAAAAGTTACAGTATCT..ACCCAGCACAGAGTTTAAGTTT 247
114 roAspLeuSerGluHisIleLysAsnPro...GlyPheAsnLeu 127
246 CAGACGTTAAAAAAATCTGTCCCGGCCAGGGGAGCTGAACATC 203
```

seq\_name: gb\_est1:AW036063

seq\_documentation\_block:

LOCUS AW036063 480 bp mRNA linear EST 18-MAY-2001  
DEFINITION EST278077 tomato seed, TAMU Lycopersicon esculentum cDNA clone  
cLEELP5, mRNA sequence.

ACCESSION AW036063  
VERSION AW036063.1 GI:5894905  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.

REFERENCE 1 (bases 1 to 480)  
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,  
'Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman  
,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley  
,S.D. and Giovannoni,J.  
Generation of ESTs from tomato seed tissue  
unpublished (1999)

TITLE JOURNAL  
COMMENT Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: http://www.genome.clemson.edu/orders/index.html  
5 prime sequence.

#### FEATURES

source Location/Qualifiers  
1..480  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEELP5"  
/clone\_lib="tomato seed, TAMU"  
/tissue\_type="seeds"  
/dev\_stage="quiescent seed"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; cLEF - Tomato Seed EST Library. Directionally cloned  
cDNAs inserted into pBluescript SK(-) at 5' end with  
EcoRI and 3' end with XhoI site."

BASE COUNT 175 a 79 c 64 g 162 t

ORIGIN

#### alignment\_scores:

Quality:	75.00	Length:	80
Ratio:	1.705	Gaps:	5
Percent Similarity:	55.000	Percent Identity:	31.250

#### alignment\_block:

US-09-728-670-10 x AW036063 ..

Align seg 1/1 to: AW036063 from: 1 to: 480

```
69 AspAlaThrAlaTyrLysGluPheArgValValGluLeuAspProSerAl 85
||||| |||:::|:::|:::|:::|:::|:::|:::|:::|:::|
138 GATGAAGACGCTTATACACTATTTCAGAAATCATCTCCGAGATATT 175
85 aLysIleGlu.....ValThrTyrT 92
176 AAATATATGAAGACGCTTATACACTATTTCAGAAATCATCTCCGAGATATT 225
92 tyrAspLysAsnLysLysLysGlu.....GluThrLysSerPheProIle 106
226 ACGATGAATAATGAAGAAGAGCTTCAAAAGAGTAAATTCGTTTCATATG 275
107 ThrGluLysGlyPheValValProAsp.LeuSerGluHisIleLysAsn 123
276 CTACCAAAAAGGAGCTTAATTCCTCTCTTGACCATCAGCTAGACACA. 324
123 roGlyPheAsnLeuIleThrLysValValIleGluLys 135
325 .....ATGAGATGTGACTTATCTAATTATTCCTTAA 357
```

seq\_name: gb\_est1:AU052970

seq\_documentation\_block:

LOCUS AU052970 553 bp mRNA linear EST 28-APR-1999  
DEFINITION AU052970 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
discoideum cDNA clone SLF410, mRNA sequence.  
ACCESSION AU052970  
VERSION AU052970  
KEYWORDS AU052970.1 GI:4701453  
SOURCE EST.  
ORGANISM Dictyostelium discoideum.  
REFERENCE Dictyostelium discoideum.  
AUTHORS Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
1 (bases 1 to 553)  
Moriya, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,  
Yoshino, R., Mitera, B.N., Pl.M., Sato, T., Takemoto, K., Yasukawa, H.,  
Williams, J., Maeda, M., Takeuchi, I., Ochial, H. and Tanaka, Y.  
Developmental cDNA in Dictyostelium discoideum  
Unpublished (1998)  
JOURNAL Contact: Hideko Urushihara  
COMMENT Institute of Biological Sciences  
University of Tsukuba  
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
Email: d402huesakura.cc.tsukuba.ac.jp  
PROJECT - Dictyostelium discoideum cDNA project in Japan.

FEATURES  
source  
1. 553  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="SLF410"  
/clone\_lib="Dictyostelium discoideum SL (H.Urushihara)"  
/dev\_stage="slug"

BASE COUNT 289 a 105 c 24 g 135 t  
ORIGIN

alignment\_scores:  
Quality: 75.00 Length: 85  
Ratio: 1.596 Gaps: 3  
Percent Similarity: 55.294 Percent Identity: 25.882

alignment\_block:  
US-09-728-670-10 x AU052970 ..  
Align seg 1/1 to: AU052970 from: 1 to: 553

```
48 ProileLysProGlyThrThrLeuThrLysGluLysIleGluTyrTyrVa 64
||||:||||| ||| ||||| ||||| |||||
178 CCAGTTCACCAATTAATTTATTAACAAAAAGAAA..... 213
64 IGlutPalaLeuAspAlaThrAlaTyrLysGluPheArgValGluL 81
||||:||||| ||||| ||||| ||||| |||||
214 .....GCAGCATTACAGACACACAACTACACACAC 244
81 euAspProSerAlaLysIleGluValThrTyrTyrAspLys..AsnLys 96
|||:||||| ||||| ||||| ||||| |||||
245 AATCACAACAATAATTTATTAATCAACTTTAATGATCAATATATAA 294
97 LysLysGluGluThrLysSerPheProIleThrGluLysGlyPheVal.. 112
|||:||||| ||||| ||||| ||||| |||||
295 CAAAAGAGTCAACCATCATATCCAACTCTGAAAGTTATATAA 344
113 .....ValProAspLeuSerGluHisI 120
||||:||||| ||||| ||||| ||||| |||||
345 TACAAGATTAAATTGCAAAAATTCATGTTCAAGATTATCTAATCAT 394
120 IeLys 121
|||:|||||
395 TTAAA 399
```

seq\_name: gb\_est2:BM474008  
seq\_documentation\_block:  
LOCUS BM474008 1082 bp mRNA linear EST 05-FEB-2002  
DEFINITION AGENCOURT-6489861 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:5521446  
5', mRNA sequence.

ACCESSION BM474008  
VERSION BM474008.1 GI:18523050  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
1 (bases 1 to 1082)  
NIH-MGC <http://mgi.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaphs-remail.nih.gov](mailto:cgaphs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM12187 row: b column: 07  
High quality sequence stop: 333.

FEATURES  
source  
1. 1082  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5521446"  
/clone\_lib="NIH\_MGC\_71"  
/tissue\_type="leiomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2.1 kb."

BASE COUNT 333 a 336 c 180 g 231 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 75.00 Length: 113  
Ratio: 1.364 Gaps: 4  
Percent Similarity: 48.673 Percent Identity: 23.009

alignment\_block:  
US-09-728-670-10 x BM474008 ..  
Align seg 1/1 to: BM474008 from: 1 to: 1082

```
4 PheAspLysGlyLysTyrLysLys..... 11
||||:||||| ||||| ||||| ||||| |||||
375 TTTCCTCCCGAGGGGAATTTAAATAAAAAAACACTTTTTCCTCCCT 424
12 .....GlyAspAspAlaSerTyrPheGluProThrGlyProT 24
||||:||||| ||||| ||||| ||||| |||||
425 TCCTTGGGGGGGGGGGGGGCCACACCCCTCCACCCCTCCT 474
24 yrLeuMetValAsnValThrGlyValAspSerLysGlyAsnGluLeu 40
|||:||||| ||||| ||||| ||||| |||||
475 ACACCTGCCACACACACTGCCGCCACAGAACGAAAAACATGAAGAATA 524
41 SerProHisTyrValGluPheProIleLysProGlyThrThrLeuThr 57
|||:||||| ||||| ||||| ||||| |||||
525 CACCCCGCATGAACCTCAGACGCCCACTGTCCTCCGGCCACTACTC.... 569
57 sGluLysIleGluTyrTyrValGluTrpAlaLeuAspAlaThrAlaTyr 74
||||:|||||
570 .....CTAAACGCCCG..... 581
74 ysgLupheArgValValGluLeuAspProSerAla.....LysIleGlu 88
|||:||||| ||||| ||||| ||||| |||||
582 .....CGTATCATCAACGTTCAACCGCACACACGTTACAAACTCAA 623
89 ValThrTyrTyrAspLysAsnLysLysGluGluThr 101
|||:||||| ||||| ||||| ||||| |||||
```







100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>.  
Location/Qualifiers

## FEATURES

source  
1. .522  
/organism="Lycopersicon esculentum"  
/cultivar="TA496, E6203"  
/db\_xref="taxon:4081"  
/clone="CTOS10E14"  
/tissue\_type="suspension cultures"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Suspension cultures of L.esculentum E6203 were grown  
in Murashige and Skoog based medium, supplemented with 15%  
coconut milk (filter sterilized and added after  
autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).  
Fresh medium was added every 7 days, and cultures were  
grown at 25 C, with 12hrs of light and continuous  
shaking."

BASE COUNT 155 a 103 c 99 g 165 t  
ORIGIN

## alignment\_scores:

Quality: 74.50 Length: 118  
Ratio: 1.183 Gaps: 5  
Percent Similarity: 53.390 Percent Identity: 24.576

## alignment\_block:

US-09-728-670-10 x BI206210/rev ..

Align seg 1/1 to reverse of: BI206210 from: 1 to: 522

```
38 GluLeuLeuSerProHisTyrValGluPheProIleLysProGlyThrTh 54
   :::::|||||:::||||| ||| |||:::
498 GATTTCCTGAAACCCCATACAGTA.....GGAAGCAG 467

54 rLeuThrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAspAlaT 71
   :::: ||| :::: ||| :::::|||||:::
466 GATCAAAAACCTCCAAAATGATATGACATCAACTATGCACCTCGACAGCA 417

71 hrAla.....TyrLysGluPheArg 77
   ::|||
416 GTGCAAAAATGAAACCCGTGGGAGAGATTATGATGAAGAAGTCCGGAAGA 367

78 ValValGluLeuAsp.....ProSerAlaLysIleGluValThrTy 91
   ||:::|||||::: |||:::|||||:::
366 GTGATGAGTGTAAAGACATCTGCACCTGTGTTCATTTCTACACTTCCAA 317

91 rTyrAspLysAsnLysLysLysGluGluThrLysSerPhePro..... 105
   ::|||:::|||||::: |||:::
316 TTGGGATTAAGTAAGGAAAGGTGATTTATGTATCCGCCCTCATG 267

106 .....IleThrGluLysGlyPheValValProAspLeuSerGluHis... 119
   ::::: ||| ::::: ||| |||||
266 CAGCATTTGGCTTTGGAAACTCTAGTATTTCCAGACGCTGTGAATCACCCCT 217

120 .....IleLysAsnProGlyPheAsnLeuIleThrLysVal 131
   ||| ||||| ||| ::::: |||
216 AATTTTCATCTCAATGTGTAATCCGGAGAGAGATACGTCATTCAGT 167

131 lVal 132
   |||
166 GTTG 163
```

seq\_name: gb\_est2:BI206708

seq\_documentation\_block:

LOCUS BI206708 522 bp mRNA linear EST 11-JUL-2001  
DEFINITION EST524748 cTOS Lycopersicon esculentum cDNA clone CTOS11F14 5' end,  
mRNA sequence.  
ACCESSION BI206708  
VERSION BI206708.1 GI:14684432

## KEYWORDS

EST.  
SOURCE

## ORGANISM

tomato.  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.

## REFERENCE

1 (bases 1 to 522)

## AUTHORS

van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R.,  
Roning,C. and Tanksley,S.

## TITLE

Generation of ESTs from Tomato Suspension Cultures

## JOURNAL

Unpublished (2001)

## COMMENT

Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>.  
Location/Qualifiers

## FEATURES

## source

1. .522  
/organism="Lycopersicon esculentum"  
/cultivar="TA496, E6203"  
/db\_xref="taxon:4081"  
/clone="CTOS11F14"  
/clone\_1lb="CTOS"  
/tissue\_type="suspension cultures"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Suspension cultures of L.esculentum E6203 were grown  
in Murashige and Skoog based medium, supplemented with 15%  
coconut milk (filter sterilized and added after  
autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).  
Fresh medium was added every 7 days, and cultures were  
grown at 25 C, with 12hrs of light and continuous  
shaking."

BASE COUNT 155 a 103 c 99 g 165 t  
ORIGIN

## alignment\_scores:

Quality: 74.50 Length: 118  
Ratio: 1.183 Gaps: 5  
Percent Similarity: 53.390 Percent Identity: 24.576

## alignment\_block:

US-09-728-670-10 x BI206708/rev ..

Align seg 1/1 to reverse of: BI206708 from: 1 to: 522

```
38 GluLeuLeuSerProHisTyrValGluPheProIleLysProGlyThrTh 54
   :::::|||||:::||||| ||| |||:::
498 GATTTCCTGAAACCCCATACAGTA.....GGAAGCAG 467

54 rLeuThrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAspAlaT 71
   :::: ||| :::: ||| :::::|||||:::
466 GATCAAAAACCTCCAAAATGATATGACATCAACTATGCACCTCGACAGCA 417

71 hrAla.....TyrLysGluPheArg 77
   ::|||
416 GTGCAAAAATGAAACCCGTGGGAGAGATTATGATGAAGAAGTCCGGAAGA 367

78 ValValGluLeuAsp.....ProSerAlaLysIleGluValThrTy 91
   |||:::|||||::: |||:::|||||:::
366 GTGATGAGTGTAAAGACATCTGCACCTGTGTTCATTTCTACACTTCCAA 317

91 rTyrAspLysAsnLysLysLysGluGluThrLysSerPhePro..... 105
   ::|||:::|||||::: |||:::
316 TTGGGATTAAGTAAGGAAAGGTGATTTATGTATCCGCCCTCATG 267

106 .....IleThrGluLysGlyPheValValProAspLeuSerGluHis... 119
   ::::: ||| ::::: ||| |||||
266 CAGCATTTGGCTTTGGAAACTCTAGTATTTCCAGACGCTGTGAATCACCCCT 217

120 .....IleLysAsnProGlyPheAsnLeuIleThrLysVal 131
```

```

216 AATTTTCATCTCAATGTGTAATCCGGAGAGAGATACGTCATTCAGT 167
131 lval 132
166 GTTG 163

seq_name: gb_est2:BI207927

seq_documentation_block:
LOCUS      BI207927          522 bp    mRNA    linear    EST 11-JUL-2001
DEFINITION EST525967 cTOS Lycopersicon esculentum cDNA clone cTOS15D23 5' end,
ACCESSION  BI207927
VERSION     BI207927.1  GI:14685651
KEYWORDS   EST.
SOURCE      tomato.
ORGANISM    Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1 (bases 1 to 522)
AUTHORS    van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R.,
            Ronning,C. and Tanksley,S.
TITLE       Generation of ESTs from Tomato Suspension Cultures
JOURNAL     Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.

FEATURES
    source          1..522
                    /organism="Lycopersicon esculentum"
                    /cultivar="TA496, E6203"
                    /db_xref="taxon:4081"
                    /clone="cTOS15D23"
                    /clone_1lb="cTOS"
                    /tissue_type="suspension cultures"
                    /lab_host="SOLR"
                    /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                    XhoI; Suspension cultures of L.esculentum E6203 were grown
                    in Murashige and Skoog based medium, supplemented with 15%
                    coconut milk (filter sterilized and added after
                    autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
                    Fresh medium was added every 7 days, and cultures were
                    grown at 25 C, with 12hrs of light and continuous
                    shaking."

BASE COUNT      156 a      103 c      98 g      165 t
ORIGIN
alignment_scores:
    Quality:      74.50      Length:      118
    Ratio:        1.183      Gaps:      5
Percent Similarity: 53.390      Percent Identity: 24.576

alignment_block:
US-09-728-670-10 x BI207927/rev ..

Align seg 1/1 to reverse of: BI207927 from: 1 to: 522

38 GluLeuLeuSerProHisTyrValGluPheProIleLysProGlyThrTh 54
   ::::::::::::::::::::
498 GATTTCCTGAACCCCATACAGTA.....GGAAGCAG 467
   ::::::::::::::::::::

54 rleuthrlysglulysileglutrytyrvalglutrpalaaleuaspalat 71
   :::: ||| :::: ||| ::::::::::::::::::::
466 GATCAAAAACCTCCAAAATGATATGACATCAACTATGCACTCGACAGCA 417
   :::: ||| :::: ||| ::::::::::::::::::::

71 hrAla.....TyrLysGluPhearg 77
   :::: ||| :::: ||| ::::::::::::::::::::
```

```

416 GTGCAAAAATGAAACCCGTGGGAGAGATTATGATAGAGTCCGGAAGA 367
78 valAlaGluLeuasp.....ProSerAlaLysilegluValThrTy 91
   ||:::::::::::::::::: ||::::::::::::::::::
366 GTGATGAGTGTAAAGCATCTGCACCTGGTGTCAATTCTACACTTCCA 317
91 rTyAspLysAsnLysLysLysGluGluThrLysSerPhePro..... 105
   :::::::::::::::::::: ::::
316 TTGGGATATAAAGTAAAAAAGGAAAGGTGATTTATGTATCCGCCTCATG 267
106 .....lIeThrGluLysGlyPheValAlaProAspLeuSerGluHis... 119
   :::::::::: :::: ||| ||| ||| ||| ||| ||| ||| |||
266 CAGCATGTGCTTGGAAACTCTAGTATTTCCAGACGCTGTGAATCACCT 217
120 .....lIeLysAsnProGlyPheAsnLeuIleThrLysVa 131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
216 AATTTCCATCTCAATGTGTAATCCGGAGAGAGATACGTCATTCAGT 167
131 lval 132
166 GTTG 163
```

seq\_name: gb\_est2:BI211112

seq\_documentation\_block:

```
LOCUS      BI211112          522 bp    mRNA    linear    EST 11-JUL-2001
DEFINITION EST529152 cTOS Lycopersicon esculentum cDNA clone cTOS25A10 5' end,
ACCESSION  BI211112
VERSION     BI211112.1  GI:14688836
KEYWORDS   EST.
SOURCE      tomato.
ORGANISM    Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
```

```
REFERENCE   1 (bases 1 to 522)
AUTHORS    van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R.,
            Ronning,C. and Tanksley,S.
TITLE       Generation of ESTs from Tomato Suspension Cultures
JOURNAL     Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
```

```
FEATURES
    source          1..522
                    /organism="Lycopersicon esculentum"
                    /cultivar="TA496, E6203"
                    /db_xref="taxon:4081"
                    /clone="cTOS25A10"
                    /clone_1lb="cTOS"
                    /tissue_type="suspension cultures"
                    /lab_host="SOLR"
                    /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                    XhoI; Suspension cultures of L.esculentum E6203 were grown
                    in Murashige and Skoog based medium, supplemented with 15%
                    coconut milk (filter sterilized and added after
                    autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
                    Fresh medium was added every 7 days, and cultures were
                    grown at 25 C, with 12hrs of light and continuous
                    shaking."
```

```
BASE COUNT      155 a      103 c      99 g      165 t
ORIGIN
```

```
alignment_scores:
    Quality:      74.50      Length:      118
    Ratio:        1.183      Gaps:      5
Percent Similarity: 53.390      Percent Identity: 24.576
```

alignment\_block:

US-09-728-670-10 x BI211112/rev ..

Align seg 1/1 to reverse of: BI211112 from: 1 to: 522

```
38 GluLeuSerProHisTyrValGluPheProIleLysProGlyThrTh 54
   ::::::::::::::::::::
498 GATTTCCTGAACCCCATACAGTA.....GGAAGCAG 467
54 rLeuThrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAspAla 71
   ::::::::::::::::::::
466 GATCAAAAACCTCAAAATGATATGACATCACTATGCACCTCGACAGCA 417
71 hrAla.....TyrLysGluPheArg 77
   ::::::::::::::::::::
416 GTGCAAAAATGAACCCGCGGAGAGTTTATGATAGAAAGTCCGGAAGA 367
78 ValValGluLeuAsp.....ProSerAlaLysIleGluValThrTy 91
   ::::::::::::::::::::
366 GTGATGAGTGTAAAGCATCTGCACCTGTGTCAATTCTACACTCCAA 317
91 rTyrAspLysAsnLysLysLysGluGluThrLysSerPhePro..... 105
   ::::::::::::::::::::
316 TTGGGATAAAGTAAAGAAAGGAAAGGTGATTTATGTATCCGCCATG 267
106 .....IleThrGluLysGlyPheValValProAspLeuSerGluHis... 119
   ::::::::::::::::::::
266 CAGCATGTGCTTGGAAACTCTAGTATTTCCAGACGCTGTGAATCACCT 217
120 .....IleLysAsnProGlyPheAsnLeuIleThrLysVa 131
   ::::::::::::::::::::
216 AATTTCATCCTCAATGTGATATCCGGAGAGAGATACGTCCATTCAGT 167
131 lVal 132
   ::::::::::::::::::::
166 GTTG 163
```

seq\_name: gb\_est2:BI955134

seq\_documentation\_block:

LOCUS BI955134 679 bp mRNA linear EST 19-OCT-2001  
DEFINITION HVSMEM0021119f Hordeum vulgare green seedling EST library  
HVCDA0014 (Blumeria infected) Hordeum vulgare cDNA clone  
HVSMEM0021119f, mRNA sequence.

ACCESSION BI955134  
VERSION BI955134.1 GI:16301210

KEYWORDS

EST.

SOURCE

barley.  
Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae

; Triticeae; Hordeum.  
1 (bases 1 to 679)

REFERENCE

Wing, R., Close, T.J., Kleinbols, A., Wise, R., Chin, A., Begum, D.,  
Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons

, J., Oates, R. and Main, D.  
Development of a genetically and physically anchored EST resource  
for barley genomics: Blumeria infected Morex (compatible) seedling  
cDNA library

Unpublished (2001)

JOURNAL

Contact: Wing RA  
Clemson University Genomics Institute

Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288  
Fax: 864 656 4293

Email: rwing@clemson.edu

Total bp bases = 479  
Seq primer: AATTAACTCTCACTAAAGG  
High quality sequence stop: 657.

FEATURES  
1..679  
/organism="Hordeum vulgare"

/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone="HVSMEM0021119f"  
HVCDA0014 (Blumeria infected)"  
/tissue\_type="green seedling leaf"  
/lab\_host="TJC121"  
/note="vector: pBluescript SK(-); Site.1: EcoRI; Site.2:  
XhoI; Morex (mla) plants were greenhouse grown in the R  
Wise lab at Iowa State University, Ames, IA; 7 day old  
green seedlings were infected with isolate 5874 of  
Blumeria graminis f. sp. hordei, and leaves were harvested  
24, 48 and 72 hr post-inoculation and snap frozen (Wise).  
In the TJC121 lab at the University of California,  
Riverside, total RNA was prepared from each sample pool,  
equal quantities of all three RNA pools were combined,  
poly(A) RNA was purified from the mixture, one primary  
unamplified cDNA library was made, and 1 million pfu were  
in vivo excised to give pBluescript SK(-) cDNA phagemids  
(Chin). Phagemids were plated and picked at the Clemson  
University Genomics Institute (CUGI) (Begum, Palmer,  
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA  
sequencing and sequence analysis were performed at CUGI  
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).  
The sequence has been trimmed to remove vector sequence  
and contains a minimum of 100 bases of phred value 20 or  
above. For more details on library preparation and  
sequence analysis see

http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders also  
see Close TJ, Wing R, Kleinbols A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT 188 a 121 c 178 g 189 t 3 others  
ORIGIN

alignment\_scores:  
Quality: 74.50 Length: 99  
Ratio: 1.552 Gaps: 3  
Percent Similarity: 48.485 Percent Identity: 25.253

alignment\_block:  
US-09-728-670-10 x BI955134 ..

Align seg 1/1 to: BI955134 from: 1 to: 679

```
33 AspSerLysGlyAsnGluLeuSer..... 41
   ::::::::::::::::::::
170 GACTCTGAAGACGTAACCACTTGATGAGCGGTTGAGGTCACGGTAT 219
42 .ProHisTyrValGluPheProIleLysProGlyThrThrLeuThrLysG 58
   ::::::::::::::::::::
220 TCCACACCTTGTTCCTTGATGCAAAACTGTGAAGTCTTACTGATG 269
58 lLysIleGluTyrTyrValGluTrpAlaLeuAspAlaThrAlaTyrLys 74
   ::::::::::::::::::::
270 AAGAGTTGAGTTGTGAAGTGAATATGTATAGAACCTTATCCTTTACA 319
75 GluPheArgValValGluLeuAspProSerAlaLysIleGluValThrTy 91
   ::::::::::::::::::::
320 ACTGACAGCATCAATGAATTCGAAGCAACA..... 349
91 rTyrAspLysAsnLysLysLysGluGluThr..... 101
   ::::::::::::::::::::
350 ....GAAGAGGCAAGTAAAGATATCAAACTATTCAATAGTGTGCTGTA 395
102 .....LysSerPheProIleThrGluLysGlyPheValValPro 114
   ::::::::::::::::::::
396 CAGCAATCGTGCCTACGTAATTCAAAACGCGGAAGAGGTACCC 442
```

seq\_name: gb\_est2:BF570018



```
seq_documentation_block:                               993 bp      mRNA       linear      EST 12-DEC-2000
LOCUS           BF570018                                Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
DEFINITION      M02186015T1 NIH_MGC_45 Homo sapiens cDNA clone IMAGE:4310571 3',
                BF570018 sequence.
ACCESSION       BF570018
VERSION         BF570018.1   GI:11643730
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE       NIH-MGC http://mgc.ncl.nih.gov/.
AUTHORS         National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE           Unpublished (1999)
JOURNAL         Contact: Robert Strausberg, Ph.D.
COMMENT         Email: cgabbs@email.nih.gov
                Tissue Procurement: Linehan
                CDNA Library Preparation: Ling Hong/Rubin Laboratory
                CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: L1CM1185 row: m column: 04
                High quality sequence start: 14
                High quality sequence stop: 714.
FEATURES        Location/Qualifiers
                1..993
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:4310571"
                    /clone_lib="NIH_MGC_45"
                    /tissue_type="renal carcinoma (ascites)"
                    /lab_host="DH10B (phage-resistant)"
                    /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
                    EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                    into EcoRI/XhoI sites using the following 5' adaptor:
                    GGCACGAG(G). Library constructed by Ling Hong in the
                    laboratory of Gerald M. Rubin (University of California,
                    Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                    Superscript II RT (Life Technologies). Note: this is a
                    NIH_MGC library."
BASE COUNT     170 a      293 c      256 g      274 t
ORIGIN
alignment_scores:
    Quality:      74.50              length:      114
    Ratio:        1.284              Gaps:        7
Percent Similarity: 50.877      Percent Identity: 30.702
alignment_block:
US-09-728-670-10 x BF570018/rev ..
Align seg 1/1 to reverse of: BF570018 from: 1 to: 993
20 ProThrGlyProTyrLeuMetValAsnValThr..... 30
||| ||||| ||:::||||: |||
796 CCTGGGGGACCACCAACGCTGTGGCAGCGCAAACCTCCAGGTGAATGAGACT 747
|||::: ||| ||||| ::|||::|||::: ::|||
31 GlyValAspSerLysGlyAsnGluLeuLeuSerProHisTyrValGluP 47
|||::: ||| ||||| ::|||::|||::: ::|||
746 GGGAACCGCTCTATGGGGAACGGCGTTCTGGAACCCAATGCTGTAAT 697
47 heProilellysProGlyThrThrLeuThrLysGlu.LysIleGluTyrTy 63
|| | ||| ||| |||::||| |||
696 TTCAGACCAAG.....ACTGAGGAAGAAG..... 673
63 rValGIuTrPalaleuAspaLaThrAlaTyrLysGluPheArgValVaIg 80
||| ||||| ||| ::|
672 .....AGACCTTGATCGGACCGCTTTGGGGG.....G 648
```

```

80 luleAspproSerAlalysileGluValThrTyrTrypasplysAnlys 96
   |||||::||||| :: ||| :: :|||:::|||||:
647 AACTGGAAACCATCTGATGA...GAATCCTCAGAAGAAAGGAAAGGAA 601
    ::::::::::||| ||| ||| |:::::
        GAAAGTGATGAACAACCACCA.....GATGAGACAGCGCTTTATTAC 560
113 lProaspLeuserglUHisIleLysAsnProglyPheasn 126
   :|||      |||       |||         |||||:::
559 CCCTGCAGACAGGTGCCCTTAACACTCCTCGAGGCTTTTCA 520

seq_name: gb_est1:AW036324

seq_documentation_block:
LOCUS      AW036324                  483 bp     mRNA          linear   EST_18-MAY-2001
DEFINITION EST278449 tomato seed, TAMU Lycopersicon esculentum cDNA clone
           CLEEIJ4, mRNA sequence.
ACCESSION  AW036324
VERSION    AW036324.1 GI:5895078
KEYWORDS   EST.
SOURCE      tomato.
ORGANISM    Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1 (bases 1 to 483)
            Alcalá,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
            Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
            ,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
            ,S.D. and Giovannonni,J.
            Generation of ESTs from tomato seed tissue
            Unpublished (1999)
TITLE       Contact: CGI
JOURNAL     Clemson University Genomics Institute
COMMENT     Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.

FEATURES
             location/Qualifiers
                source              1..483
                                     /organism="Lycopersicon esculentum"
                                     /cultivar="TA496"
                                     /db_xref="taxon:4081"
                                     /clone="CLEEIJ4"
                                     /clone_lib="tomato seed, TAMU"
                                     /tissue_type="seeds"
                                     /dev_stage="quiescent seed"
                                     /lab_host="XLI-Blue MRF"
                                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                                     XhoI; cleE - Tomato Seed EST Library. Directionally cloned
                                     cDNAs inserted into pBluescript SK(-) at 5' end with
                                     EcoRI and 3' end with XhoI site."

BASE COUNT      179 a               77 c               64 g               163 t
ORIGIN

alignment_scores:
Quality:      74.00                      Length:      80
Ratio:        1.762                     Gaps:        5
Percent Simlarity:  52.500                 Percent Identity: 32.500

alignment_block:
US-09-728-670-10 x AW036324 ..

Align seg 1/1 to: AW036324 from: 1 to: 483

69 AspaLatHraAtYrlysgLuPheargVaIvalGluleAAspproSerAl 85
   |||      |||||      |||      :::|||||
138 GATGAAGACGCTTATACACTATTTCAC.....GAACCATCTCC 175
```



Align seg 1/1 to: BI391530 from: 1 to: 613

```
20 ProThrGlyProTyrLeuMetValAsnValThrGlyValAspSerLysG1 36
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35 CCGCCGGGGCGGCTGCTGCTGGGGCGGCTGCTGGGTCTCTCTCA..... 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
36 yAsnGluLeuLeuSerProHisTyr..... 44
   ::|||::: |||
80 ...GCTGTGCTGAGGTGCACGCTGGAGACGGGCTGTCTACTCAGTGG 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
45 ..ValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIle 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 AAGGGAGGAGCGCGTGTGCTGCGCCCTGTACACACAGCAGCTCAGAGAAG 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61 GluTyrTyrValGluTyrPalaLeuAspAlaThrAlaTyrLysGluPheAr 77
   ::|||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
176 AGCCCTACATCAGCTGATGCTGAAAAAGCAGCTGCTAAGCAATTCCA 225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
77 gValValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspL 94
   ::|||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
226 GATC.....CTGACATACCTGAGCG 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94 ySAsnLysLysLysGluGluThrLysSerPheProIleThrGluLysGly 110
   ::|||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
246 GGGAGGTGAAGTGAGGAGACAGAGCTGAAGCCC.....CGTGTGGGG 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
111 pheValValProAspLeuSerGluHisIle 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
290 TTCCTGCACCCCGTCTCACCACACATC 319
seq_name: gb_est2:BI065721
```

seq\_documentation\_block:

LOCUS BI065721 654 bp mRNA linear EST 15-JUN-2001  
DEFINITION pgfin.pk005.k14 normalized chicken fat cDNA library Gallus gallus  
CDNA clone pgfin.pk005.k14 5' similar to g110048456  
reflmp\_065264.1 cortical thymocyte receptor (X. laevis CTX) like;  
CTM gene [Mus musculus] gb|AADI7524.11 (AF061024) CTM [Mus musculus]  
]G, mRNA sequence.  
ACCESSION BI065721  
VERSION BI065721.1 GI:14473243  
KEYWORDS EST.  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 654)  
AUTHORS Cogburn,L.A., Morgan,R.W. and Burnside,J.  
TITLE Chicken ESTs from fat  
JOURNAL Unpublished (2001)  
COMMENT Contact: Larry A. Cogburn  
University of Delaware  
Townsend Hall, Newark, DE 19717, USA  
Tel: 302-831-1335  
Fax: 302-831-2822  
Email: cogburn@udel.edu, www.chickest.udel.edu.  
FEATURES  
source  
1..654  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/clone="pgfin.pk005.k14"  
/clone\_lib="normalized chicken fat cDNA library"  
/sex="Male and female"  
/tissue\_type="fat"  
/lab\_host="E.coli EMDH10B"  
/note="Vector: pSPORT1"

BASE COUNT 133 a 206 c 193 g 111 t 11 others  
ORIGIN

alignment\_scores:

Quality: 74.00 Length: 110  
Ratio: 1.370 Gaps: 4  
Percent Similarity: 49.091 Percent Identity: 26.364

alignment\_block:

US-09-728-670-10 x BI065721 ..  
Align seg 1/1 to: BI065721 from: 1 to: 654

```
20 ProThrGlyProTyrLeuMetValAsnValThrGlyValAspSerLysG1 36
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4 CCGCCGGGGCGGCTGCTGCTGGGGCGGCTGCTGGGTCTCTCTCA..... 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
36 yAsnGluLeuLeuSerProHisTyr..... 44
   ::|||::: |||
49 ...GCTGTGCTGAGGTGCACGTCGGAACGGGCTGTCTACTCAGTGG 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
45 ..ValGluPheProIleLysProGlyThrThrLeuThrLysGluLysIle 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 AAGGGAGGAGACCGCTGCTGCGCCGCTGTACACACAGCAGCTCAGAGAAG 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61 GluTyrTyrValGluTyrPalaLeuAspAlaThrAlaTyrLysGluPheAr 77
   ::|||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 AGCCCTACATCAGCTGATGCTGAAAAAGCAGCTGCTAAGCAATTCCA 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
77 gValValGluLeuAspProSerAlaLysIleGluValThrTyrTyrAspL 94
   ::|||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 GATC.....CTGACATACCTGAGCG 214
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94 ySAsnLysLysLysGluGluThrLysSerPheProIleThrGluLysGly 110
   ::|||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
215 GGGAGGTGAAGTGAGGAGACAGAGCTGAAGCCC.....CGTGTGGGG 258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
111 pheValValProAspLeuSerGluHisIle 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
259 TTCCTGCACCCCGTCTCACCACACATC 288
seq_name: gb_est2:BI929102
```

seq\_documentation\_block:

LOCUS BI929102 718 bp mRNA linear EST 18-OCT-2001  
DEFINITION EST548991 tomato flower, 3 - 8 mm buds Lycopersicon esculentum CDNA  
clone CTOB27B225' end, mRNA sequence.  
ACCESSION BI929102  
VERSION BI929102.1 GI:16242002  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
REFERENCE 1 (bases 1 to 718)  
AUTHORS van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,  
Utterback,T., Van Aken,S., Ronning,C.M., Nieman,W., Fraser,C.M.,  
Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.  
TITLE Generation of ESTs from tomato flower tissue, 3-8 mm buds (2001)  
JOURNAL Unpublished (2001)  
COMMENT Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
This clone is available through the Clemson University Genomics  
Institute  
Seq primer: T3.  
FEATURES  
source  
1..718  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CTOB27B22"  
/clone\_lib="tomato flower, 3 - 8 mm buds"

```

/tissue_type="flower"
/dev_stage="3-8mm buds"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Cornell University; sequencing: The
Institute for Genomic Research. Flower buds and flowers
were taken from greenhouse plants (4-8 wks old, TA496).
They were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."

```

```

BASE COUNT      223 a      107 c      202 g      186 t
ORIGIN

alignment_scores:
    Quality:      74.00      Length:      105
    Ratio:        1.298      Gaps:        6
    Percent Similarity: 54.286      Percent Identity: 28.571

```

```

alignment_block:
US-09-728-670-10 x BI929102 ..

```

```

Align seg 1/1 to: BI929102 from: 1 to: 718

```

```

18 PheGluProThrGlyProTyrLeuMetValAsnValThrGlyValAspSe 34
   :::::::::::::::::::: ::::::::::::::::::::
123 TGGGATCCCACTGGAGAGGTATGTAGCAGACGCTGTACTTCATTCATGA 172
   :::::::::::::::::::: ::::::::::::::::::::
34 rLySGLyAsn.....G 38
   ::::::::::::::::::::
173 GATGGAATAATGGATTCAACATCTGCTTTCAACGGAAGCTGCTGATA 222
   ::::::::::::::::::::
38 LuLeuLeuSerProHisTyrValGluPhe.....ProLeuLysPro 51
   :::::::::::::::::::: ::::::::::::::::::::
223 GGATGACTCAAGGATCATTTTTCAGTTTGTGGCGCCT..AGGCCA 269
   :::::::::::::::::::: ::::::::::::::::::::
52 GlyThrThrLeuThrLysGluLysIleGluTyrTyrValGluTrpAlaLe 68
   :::::::::::::::::::: ::::::::::::::::::::
270 CCATCCCTCTTAAGTAAGAGAGAAAGAGAA.....GAGATTGCTTAA 310
   ::::::::::::::::::::
68 uAspAlaThrAlaTyr..LysGluPheArgValValGluLeuAspProS 84
   :::::::::::::::::::: ::::::::::::::::::::
311 GAACCTGAAGAAGTACAGCAAGAAGTATGAAGCAGAGATCAGATGTTT 360
   :::::::::::::::::::: ::::::::::::::::::::
84 eAlaAlaLysIleGluValThrTyrTyrAspLysAsnLysLys..LysGlu 99
   :::::::::::::::::::: ::::::::::::::::::::
361 CATGCTGTGTGAGCGAGCAGACAGCCGTGAGAAAGCAGAAAGCTGAAGA 410
   ::::::::::::::::::::
100 GluThrLysSerPhe 104
   ::::::::::::::::::::
411 GAATGGGAGCATGC 425

```

```

seq_name: gb_gss:AZ671442

```

```

seq_documentation_block:

```

```

LOCUS      AZ671442      830 bp      DNA      linear      GSS 14-DEC-2000
DEFINITION Entamoeba histolytica sheared DNA Entamoeba histolytica
              genomic, DNA sequence.
ACCESSION  AZ671442
VERSION    AZ671442.1 GI:11808588
KEYWORDS   GSS.
SOURCE      Entamoeba histolytica.
            Entamoeba histolytica.
ORGANISM   Entamoeba histolytica.
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE  1 (bases 1 to 830)
            Loftus, B., Van Aken, S. and Fraser, C.
            Determination of clone end sequences from Entamoeba histolytica
            HMI:IMSS sheared DNA library
            Unpublished (2000)

```

```

JOURNAL
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543

```

```

Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 16
High quality sequence stop: 818.
Location/Qualifiers
1. .830
   /organism="Entamoeba histolytica"
   /strain="HMI:IMSS"
   /db_xref="taxon:5759"
   /clone_lib="Entamoeba histolytica sheared DNA"
   /note="Vector: pROSI; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + l method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

```

BASE COUNT      250 a      129 c      61 g      390 t
ORIGIN

```

```

alignment_scores:
    Quality:      73.50      Length:      99
    Ratio:        1.470      Gaps:        4
    Percent Similarity: 50.505      Percent Identity: 24.242

```

```

alignment_block:
US-09-728-670-10 x AZ671442/rev ..

```

```

Align seg 1/1 to reverse of: AZ671442 from: 1 to: 830

```

```

56 ThrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAspAla..... 70
   :::::::::::::::::::: ::::::::::::::::::::
405 TCAGAGAAAGAAATAGAAAGATTATTAATGATGGCGCAAAAGAAATTAGAATT 356
   ::::::::::::::::::::
71 .....ThrAlaTyrLysGluPhe 77
   ::::::::::::::::::::
355 AATTGATGAAAAAGAAATTAAGAAATCATTCAGTGAATACAAAGAAATTA 306
   ::::::::::::::::::::
77 rGValValGluLeuAspProSerAlaLysIle..... 87
   ::::::::::::::::::::
305 GATATGAAGAAAAAGAGAAAAAGTAAATTAATAAAGAAACAGAAATCT 256
   ::::::::::::::::::::
88 GluValThrTyrTyrAspLysAsnLysLysGluGluThrLysSerPhe 104
   ::::::::::::::::::::
255 GATATTAACATAT.....GAAAAAAGACATACAGAAAGATGTAT 218
   ::::::::::::::::::::
104 eProIleThrGluLysGlyPheValValProAspLeuSerGluHisIleL 121
   ::::::::::::::::::::
217 TAAAGTAACAATGATGGA.....GAAGAAATAG 189
   ::::::::::::::::::::
121 yAsnProGlyPheAsnLeuIleThrLysValValIleGluLysLys 136
   ::::::::::::::::::::
188 AAGGTGAATGATATTATTATTAGTAAACTAATAACAGAAACAAA 142
   ::::::::::::::::::::

```

```

seq_name: gb_est1:A1164607

```

```

seq_documentation_block:

```

```

LOCUS      A1164607      355 bp      mRNA      linear      EST 03-DEC-1998
DEFINITION A065P71u Hybrid aspen plasmid library Populus tremula x Populus
              tremuloides cDNA 5', mRNA sequence.
ACCESSION  A1164607
VERSION    A1164607.1 GI:3855892
KEYWORDS   EST.

```



```

SOURCE          Populus tremula x Populus tremuloides.
ORGANISM        Populus tremula x Populus tremuloides
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE
AUTHORS         Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A., Holmberg
                ,A., Aminl,B., Bhalerao,R., Larsson,M., Villarroel,R., Van Montagu
                ,M., Sandberg,G., Olsson,O., Teerl,T.T., Boerjan,W., Gustafsson,P.,
                Uhlen,M., Sundberg,B. and Lundeberg,J.
TITLE           Gene discovery in the wood-forming tissues of poplar: Analysis of
                5,692 expressed sequence tags
JOURNAL         Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
MEDLINE         99007314
COMMENT         Contact: Sterky F
                Department of Biotechnology
                Royal Institute of Technology (KTH)
                Teknikringen 34, S-100 44 STOCKHOLM, Sweden
                Tel: +46 8 790 8287
                Fax: +46 8 24 54 52
                Email: fredrik@biochem.kth.se
                PCR Primers
                FORWARD: AAAGGGGATGCTGCTCAAGCGG
                BACKWARD: GCTCCGGCTGTATGTGTG
                Seq primer: CGTTGTAACGACGCCAG
                High quality sequence stop: 355.
                Location/Qualifiers
                1..355
                /organism="Populus tremula x Populus tremuloides"
                /db_xref="taxon:47664"
                /clone_lib="Hybrid aspen plasmid library"
                /tissue_type="Cambial region"
                /dev_stage="1.5 m actively growing tree"
                /lab_host="E.coli"
                /note="Vector: Bluescript SK; Site_1: SalI; Site_2: NotI;
                Cambial region tissues, including developing xylem, the
                meristematic cambial zone and the developing and mature
                phloem, was harvested from 1.5 m actively growing trees.
                cDNA was prepared and cloned into lambda gt22a. DNA was
                isolated and subcloned into pBluescript SK using SalI and
                NotI restriction enzymes."
BASE COUNT      118 a 79 c 87 g 68 t 3 others
ORIGIN
alignment_scores:
                Quality: 73.00          Length: 56
                Ratio: 2.281            Gaps: 1
                Percent Similarity: 57.143 Percent Identity: 30.357
alignment_block:
US-09-728-670-10 x A1164607 ..
Align seg 1/1 to: A1164607 from: 1 to: 355
2 SerSerPheaspIysGlyLysTyrLysGlyaspAspalaSerTyrph 18
||||| :::||||| ::::: ::: |||||
35 TCAGTCGAGGAGGCAAAACGAGAGGNGGAANCAAGAGATATT 84
18 egluProThrGlyPro.....TyrL 25
|||| :::||||
85 CGAGGGAATAGCTCTAAGCGTCACTTAAGCCTCAACGCAGCGATATT 134
25 eumetValasnValThrGlyValaspSerLysGlyasnGluLeuSer 41
::: ||| ::::: :::||||| :::::
135 CCATTCAATACGTTGATGCGCTCCCACTAAGGGCGATCAAAAGCTTACC 184
42 ProHisTyrValGluPhe 47
||||:|||||
185 CCAGAAATATGTTGAGTTT 202
seq_name: gb_est2:T19018

```

```

seq_documentation_block:
LOCUS      T19018                      402 bp      mRNA      linear      EST 25-SEP-1996
DEFINITION f11010t Testis 1 Homo sapiens cDNA clone f11010 5' end, mRNA
ACCESSION  T19018
VERSION    T19018
KEYWORDS   T19018.1  GI:601061
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 402)
            Pawlak, A., Toussaint, C., Levy, I., Bulle, F., Poyard, M., Barouki, R.
            and Guellaen, G.
            Characterization of a large population of mRNAs from human testis
            Genomics 26, 151-158 (1995)
            95301283
TITLE       Contact: Guellaen G
JOURNAL     Unite INSERM 99
MEDLINE     Contact INSERM 99
COMMENT     INSERM
            Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France
            Tel: (33)149813530
            Fax: (33)148980908
            Email: guellaen@infobiogen.fr
            Seq primer: T7.
FEATURES
            Location/Qualifiers
                1..402
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="f11010"
                /clone_1lb="Testis 1"
                /note="Vector: pGEM 5zf(+); Site_1: EcoRV; Site_2: NotI;
                mRNA was prepared from human testis of a 27 years old man.
                cDNA was prepared using a 15mer oligo dT anchored by two
                degenerated bases at its 3' end and containing a NotI site
                at its 5' end. The cDNA was cloned between EcoRV and NotI
                sites of pGEM 5zf(+). The 3' end is at the NotI site. The
                EcoRV site is lost during the cloning procedure. cDNA
                corresponding to abundant species were eliminated from
                this library."
BASE COUNT  130 a      66 c      97 g      102 t      7 others
ORIGIN
alignment_scores:
            Quality: 73.00      Length: 115
            Ratio: 1.237      Gaps: 5
Percent Similarity: 51.304      Percent Identity: 27.826
alignment_block:
US-09-728-670-10 x T19018      ..

Align seg 1/1      to: T19018      from: 1      to: 402

26 MetValasNValThrGlyVal..AspSerLysGlyAsnGluLeuSe 41
   ::::::::::: ||| ||| ::::::::::: ::
17 ATAGTTGACATTACTTTTGTGATGACTATAGAGGAGCGGAAAAACAG 66
   ATAGTTGACATTACTTTTGTGATGACTATAGAGGAGCGGAAAAACAG 66
41 rProHisTyrValGluPhe.....ProIleLysProGlyThrThrLeuT 56
   : ||::::: ||| ||| ::::: |||
67 GGAAGCCATATGTCAATTTGAAGAACCAAGAAATGCCCAACCAAGCCCTGT 116
   GGAAGCCATATGTCAATTTGAAGAACCAAGAAATGCCCAACCAAGCCCTGT 116
56 hTrLysGluLysIleGluTyrTyrValGluTrpAlaLeuAspAlaThrAla 72
   ||::::: |||
117 TGAACACACAGGGAAGAATTT..... 136
   TGAACACACAGGGAAGAATTT..... 136
73 TyrLysGluPheArgValValGluLeuAspProSerAlaLysIleGluVa 89
   ||| ::::::::::: ||||| ::::: |||||
137 .....GSTAAATCGATACATCGAGATATTTCCAAAGCAGAAGGAATGAGT 180
   .....GSTAAATCGATACATCGAGATATTTCCAAAGCAGAAGGAATGAGT 180
89 lThrTyrTyrAspLysAsnLysLysGluGluThrLysSerPheProI 106
   l ::::: ||| ::::::: ||||| |||||
181 TCGAACACATGTCGGTCTCTATATAAGGGAAGAAATCGCATCTTTTCTT. 229
   TCGAACACATGTCGGTCTCTATATAAGGGAAGAAATCGCATCTTTTCTT. 229

```

106 leThrgLUySGlypheValValProAspleu.....SerGluHisIle 120  
||| ||| ::::: |||::: ::|||  
230 .ACTGCTAAGTATATTAAGTCCAGAACAAATGGCTTTTGAAGAACATGAN 277  
121 LysAsnProGlypheAsnLeuIleThrLySValValIleGluLys 135  
||| |||::: ::||| |||  
278 GTAATGAGGGTATTCACACCATGACAGCTTTTGAAGTAGAGAG 322

seq\_name: gb\_est2:BF158099

seq\_documentation\_block:

LOCUS BF158099 457 bp mRNA linear EST 23-FEB-2001  
DEFINITION fl28el0.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone  
3815154 5', mRNA sequence.

ACCESSION BF158099  
VERSION BF158099.1 GI:11053303

KEYWORDS EST.  
SOURCE zebrafish.  
ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
; Cyprinidae; Danio.  
1 (bases 1 to 457)

REFERENCE  
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy  
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood  
,K., Steptoe,M., Thelsting,B., Allen,M., Bowers,Y., Person,B.,  
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,  
Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.  
and Wilson,R.

TITLE Washu Zebrafish EST Project 1998  
JOURNAL Unpublished (1998)  
COMMENT Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: zbratfish@watson.wustl.edu  
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA  
Sequencing by: Washington University Genome Sequencing Center Clone  
distribution information can be found through the I.M.A.G.E.  
Consortium/LNL, send email to: info@image.lnl.gov  
Seq primer: T3 ET from Amersham  
High quality sequence stop: 415.  
Location/Qualifiers

FEATURES  
source 1..457  
/organism="Danio rerio"  
/strain="AB"  
/db\_xref="taxon:7955"  
/clone="3815154"  
/clone\_lib="Sugano Kawakami zebrafish DRA"  
/sex="mixed (one male and one female, including  
unfertilized eggs)"  
/dev\_stage="adult"  
/lab\_host="DH10B (phage resistant)"  
/note="Vector: pME18S-FL3; site\_1: DraIII (CACTGTGTG);  
site\_2: DraIII (CACCATGTG); 1st strand cDNA was primed  
with an oligo(dT) primer [ATGTGGCCTTTTCTTTTCTTTT];  
double-stranded cDNA was ligated to a DraIII adaptor  
[TGTGGCCTACTGG], digested and cloned into distinct DraIII  
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site  
CACCATGTG). XhoI should be used to isolate the cDNA  
insert. Size selection was performed to exclude fragments  
<1.5kb. Library constructed by Dr. Sumio Sugano  
(University of Tokyo Institute of Medical Science) and  
kindly donated by Dr. Koichi Kawakami. Custom primers for  
sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end  
primer CGACCTGACGCTCGACACA."

BASE COUNT 127 a 99 c 120 g 111 t  
ORIGIN

alignment\_scores:

Quality: 73.00 Length: 110  
Ratio: 1.217 Gaps: 6  
Percent Similarity: 54.545 Percent Identity: 25.455

alignment\_block:

US-09-728-670-10 x BF158099 ..

Align seg 1/1 to: BF158099 from: 1 to: 457

24 TyrLeuMetValAsnValThrgLUyAlaspSer..... 34  
||| ::| ::::: ::|||  
150 TACAGACTGCCACAGATCGTCAAGTTGGACAGCGGAGAAATCGGTGAAGC 199  
35 ...LysGLYAsnGluLeuSerProHisTyrValGluPheProIleL 50  
::| ::| ::| ::|  
200 GTTGAAGAGAAATGACTTACTCTTGTATTCAT.....TCATGCC 237  
50 yspProGlyThrThrLeuThrLysGluLysIleGlu.....TyrTyr 63  
::| ::| ::| ::| ::| ::|  
238 GGCAGTGGACACCAATTCACGCCCATAGCCTAGAGAGGGCCATTATGTC 287  
64 ValGluTrrPalaleuAspAlaThrAlaTyrLysGlu.....PheArgVa 78  
::| ::| ::| ::| ::| ::|  
288 ATTCGACCGAAATCGAGATCCCTGTCATTATGAGGCTCAGTTAAGCT 337  
78 lValGluLeuAspProSerAlaLysIleGluValThrTyrAspLysA 95  
::| ::| ::| ::| ::| ::|  
338 GCTGAGACGACGACAGGACATTAAGAGACCTGTGACACTACTAC.....A 381  
95 snLysLysLysGluGluThrLysSerPheProIleThrgLUySGlyPhe 111  
||| ::| ::| ::| ::| ::| ::|  
382 ACAGTGTGGAGAGAGGTGGCCCAAGCTTTTCCT.....GAGCAGTGTAT 425  
112 ValValProAspleuSerGluHisIleLys 121  
||| ::| ::| ::| ::| ::| ::|  
426 GTCATGAGAGGAAATATCTTCATATTTCAG 455

seq\_name: gb\_est1:AL389514

seq\_documentation\_block:

LOCUS AL389514 527 bp mRNA linear EST 03-AUG-2000  
DEFINITION MtBC55D02R1 MtBC Medicago truncatula cDNA clone MtBC55D02 T7, mRNA  
sequence.

ACCESSION AL389514  
VERSION AL389514.1 GI:9689265

KEYWORDS EST.  
SOURCE barrel medic.

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.  
1 (bases 1 to 527)

REFERENCE  
AUTHORS Journet,E.P., Crespeau,H., van-Tulnen,D., Guzy,J., Jallion,O.,  
Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianlinazzi-Pearson

,V. and Gamas,P.  
Medicago truncatula ESTs from endomycorrhizal roots  
unpublished (2000)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de  
Biologie Molculaire des Relations Plantes-Microorganismes,  
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :  
mc-est@toulouse.inra.fr Website :  
http://sequence.toulouse.inra.fr/Mtruncatula.html).

FEATURES  
source 1..527  
/organism="Medicago truncatula"  
/cultivar="Jemalong"  
/db\_xref="taxon:3880"  
/clone="MtBC55D02"

/clone\_lib="MtBC"  
/tissue\_type="arbuscular mycorrhiza"  
/dev\_stage="harvested 3 weeks post inoculation with Glomus intraradices"  
/note="Vector: pBluescript psk; Site\_1: EcoRI; Site\_2: XhoI; M. truncatula sterilised seeds were germinated for 72h at 25 C, before transplanting into a 1/3 Epilobes soil : 2/3 catclined Terragreen mix in the presence of onion root fragments colonized by the arbuscular mycorrhizal fungus Glomus intraradices (Schenck & Smith, isolate LPA8 ). The plants were watered every day and twice a week with a modified nutrient long Ashton solution without phosphate but with a high level of nitrate. After 3 weeks RNA was extracted from whole root systems. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using Exassit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France). Note : EST may be of fungal origin."

BASE COUNT	175 a	89 c	96 g	167 t
ORIGIN				

alignment_scores:		
Quality:	73.00	Length: 94
Ratio:	1.553	Gaps: 4
Percent Similarity:	50.000	Percent Identity: 26.596

```
alignment_block:
US-09-728-670-10 x AL389514 . .
```

Align seg 1/1 to: AL389514 from: 1 to: 527

```

62 TyrTyrValGluThrPalalaleuaspalaThrAlaTyrllysGluPheargVa 78
   :: ::::|||||: |||||::: ||::|||
157 TTTTCACCTTCGTGGAGTCTCGATGCACAATAGTTGAAGAACCCTCTCCCG 206
   || | |||||::: ||::: |||||
78 lValGluLeuaspProSerAlaLysileGluVal.....ThrTyr. 91
   || | |||||::: ||::: |||||
207 TGGTGCTTTAAATCCCAAGTTCACTGATTAAGTTTACCAACGAGCACCTAAT 256
   .....TyrrAsplysAnlyslslyls 98
   |||:::|||||: |||
257 TTTCGAAAAGAAAAAAACTAAATACACTATAATAGAAAGAAAGAGAGAGC 306
   ::||| ||||| |||||
99 GluGluThrLysSerPheProIlethrGlulysGlyPheValValProas 115
   ::||| ||||| |||||
307 AAAGAACAACAANAATCTTTC..... 324
115 pleuSerglunHisIleLysasnPro..... 123
   |||::: |||
325 ....TTTGCCCATTTAAGCTTTCACATTCCTAATGTTGGTGGCATACTAG 370
124 ..GlyPheasnleuIleThrLysValValIle 133
   ::|||::: ||||| |||||
371 GAAGTTTAGACATTATGTGTGAAGCGTGGTGATT 402
```

